



# Étude des formes monogéniques de diabète de type 2 et d'obésité par le séquençage de nouvelle génération

Julien Philippe

## ► To cite this version:

Julien Philippe. Étude des formes monogéniques de diabète de type 2 et d'obésité par le séquençage de nouvelle génération. Médecine humaine et pathologie. Université du Droit et de la Santé - Lille II, 2014. Français. NNT : 2014LIL2S049 . tel-01198926

**HAL Id: tel-01198926**

**<https://theses.hal.science/tel-01198926>**

Submitted on 14 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse présentée en vue de l'obtention du titre de  
Docteur en sciences de la vie et de la santé  
*Spécialité : Génétique humaine*

Julien Philippe

# Étude des formes monogéniques de diabète de type 2 et d'obésité par le séquençage de nouvelle génération

Thèse codirigée par le Professeur Philippe Froguel et le Docteur Amélie Bonnefond

Date de soutenance : 19 décembre 2014

Jury :

Professeur Anne Vambergue	Examinatrice, Présidente du Jury
Professeur Jean-François Zagury	Rapporteur
Professeur Jacques Beckmann	Rapporteur
Docteur Amélie Bonnefond	Codirectrice de thèse

# Remerciements

Mes premiers remerciements vont aux membres du jury, les professeurs Anne Vambergue, Jean-François Zagury et Jacques Beckmann, pour avoir accepté de juger mon travail de thèse.

Je remercie mon directeur d'unité, Philippe Froguel, de m'avoir permis d'aller plus loin que ce que j'avais envisagé en faisant une thèse au laboratoire, et merci également pour les moyens financiers, scientifiques et humains qui ont été nécessaires à la réalisation de ce travail.

Je remercie ma codirectrice de thèse, Amélie Bonnefond, de m'avoir encadré au quotidien tout au long de mon travail de thèse et pour son aide dans l'écriture des articles dans lesquels je suis coauteur.

Je remercie Martine Vaxillaire de m'avoir recruté dans l'unité en tant qu'ingénieur d'études en 2010 et Aurélie Dechaume pour m'avoir formé aux techniques de séquençage durant cette première année.

Toute ma gratitude va à l'équipe NGS avec laquelle j'ai le plus travaillé ces dernières années et sans laquelle aucune des publications n'aurait été possible.

Je remercie aussi l'ensemble de l'unité UMR8199 et également EGID pour sa dynamique scientifique centrée autour des multiples conférences et événements organisés.

Je remercie nos collaborateurs, services hospitaliers et unités de recherche, qui ont contribué au succès de ce travail qui a conduit à 10 publications.

Une pensée pour les patients des cohortes et membres des familles qui ont participé à ces études.

Merci à ma famille, en particulier ma femme, mes parents et ma sœur, et à mes amis pour leur soutien. Cette thèse est dédiée à ma fille, à qui je souhaite de ne jamais devenir diabétique ni obèse !

# Résumé

Mots clés: diabète, obésité, formes monogéniques, NGS

Le diabète et l'obésité ont atteint de telles proportions dans le monde qu'on parle de pandémie. Ces deux maladies sont une cause importante de mortalité et de multiples complications. Les enjeux médicaux et financiers font de ces deux maladies un problème majeur de santé publique. Deux groupes de facteurs contribuent à ces deux maladies : l'environnement, et la génétique sur laquelle cette thèse s'appuie. Ce travail s'est focalisé sur les formes rares et monogéniques qui constituent les formes extrêmes de diabète de type 2 et d'obésité.

Ces formes sont loin d'être totalement élucidées. Mon projet s'est concentré sur l'utilisation du séquençage de nouvelle génération (NGS) pour identifier de manière plus optimale, par rapport au séquençage classique de type Sanger, des mutations dans des gènes déjà connus chez de nouveaux patients introduits dans notre cohorte dans une optique de diagnostic. Le deuxième objectif était d'utiliser les techniques de NGS pour découvrir de nouveaux loci, liés à de nouvelles voies de signalisation impliquées dans la physiopathologie du diabète et de l'obésité.

La première approche utilise une technique d'enrichissement par hybridation en phase liquide et se focalise sur 34 gènes associés à des formes monogéniques et polygéniques d'obésité. Le criblage a été réalisé sur 201 individus dans 13 familles dont la cause d'obésité est inconnue. Cette approche a mené à l'identification d'une mutation dans un gène connu de l'obésité : *PCSK1*. Cette mutation est causale, car elle conduit à un codon-stop au début de la protéine et n'est présente que chez des individus obèses. De plus, l'étude fonctionnelle a démontré une inhibition partielle de PC1/3 par la protéine tronquée et un possible impact sur la maturation et la sécrétion de l'enzyme. Cette étude a fait l'objet d'une publication dans le journal « International Journal of Obesity ».

La deuxième approche se base sur une technique d'amplification par PCR dans des microgouttelettes lipidiques développée par la société Raindance, dont le premier test vise à réidentifier les mutations causales du diabète et/ou de l'obésité chez 40 patients. Cette approche a donné des résultats satisfaisants, car pour une large majorité de patients, les mutations causales ont pu être à nouveau identifiées. Seul un patient n'a pu être reconfirmé à cause des outils bioinformatiques actuels qui restent limités dans la détection des indels complexes. Parmi les 39 patients identifiés, 3 d'entre eux sont potentiellement porteurs de plusieurs mutations causales. Cette technique pourrait être envisagée dans le domaine clinique, car elle permet une approche multigénique en fournissant un diagnostic rapide, moins coûteux et qualitativement aussi bon que le séquençage Sanger. Pour nous, la finalité de cette technique est un diagnostic clinique rapide et optimal en vue d'identifier des cas non élucidés, candidats pour le séquençage d'exome. Cette deuxième étude a été publiée dans le journal « Diabetes Care ».

La troisième approche met en jeu le séquençage de l'exome entier (WES) chez 4 individus où la famille entière s'est précédemment révélée négative pour tous les gènes connus du diabète. Cette approche a permis la découverte d'un 13<sup>e</sup> gène du MODY, *KCNJ11*, et confirme le large spectre

phénotypique qui va du diabète néonatal au MODY selon les mutations. La difficulté majeure de cette technique est le filtrage des variants en vue d'aboutir à une seule mutation causale (ou éventuellement plusieurs sur un même gène) pour identifier de nouveaux gènes du MODY. La stratégie utilisée combinait à la fois un filtrage bioinformatique, avec par exemple des filtres sur la coségrégation familiale et sur des bases de SNPs référencés, et un filtrage biologique, avec l'utilisation d'une technique de génotypage haut débit. Cette étude pionnière dans l'usage du NGS pour déterminer de nouveaux gènes du MODY a été publiée dans « PLoS ONE ».

En conclusion, ce travail a permis de tirer parti des avancées technologiques comme la capture, le séquençage ciblé de masse et le NGS pour élucider et améliorer le criblage des formes monogéniques de diabète et d'obésité. Cette amélioration de la compréhension des mécanismes moléculaires conduira peut-être au développement de meilleurs traitements comme la médecine personnalisée. On espère voir des améliorations directes pour le patient dans un futur proche, par exemple un diagnostic moléculaire plus rapide, plus sûr et plus exhaustif.

# Genetic causes of monogenic forms of diabetes and obesity

Keywords: diabetes, obesity, monogenic forms, NGS

## Abstract

Diabetes and obesity have reached such proportions worldwide we're talking about pandemic. Both diseases are a major cause of mortality and multiple complications. Medical and financial issues are for both diseases a major public health problem. Two groups of factors contribute to these two diseases: environment, and genetics on which this thesis is based. This work focused on rare and monogenic forms which are extreme forms of type 2 diabetes and obesity.

These forms are far from being fully understood. My project focused on the use of next generation sequencing (NGS) to identify more optimally, compared to conventional Sanger sequencing, mutations in already known genes among new patients in our cohort for diagnostic purposes. The second objective was to use NGS to discover new loci associated with new signaling pathways involved in the pathophysiology of diabetes and obesity.

The first approach uses a liquid-phase hybridization technique and focuses on 34 genes associated with monogenic and/or polygenic obesity. The screening was carried out on 201 people in 13 families for which the cause of obesity is unknown. This approach led to the identification of a mutation in a known gene of obesity: *PCSK1*. This mutation is causal because it leads to a stop codon at the beginning of the protein and is present only in obese individuals. Additionally, functional studies have demonstrated partial inhibition of PC1/3 by the truncated protein and the possible impact on the processing and secretion of this enzyme. This study has been published in the "International Journal of Obesity" newspaper.

The second approach is based on a PCR amplification technique in lipid microdroplets developed by Raindance. The first test is to re-identify the causal mutations of diabetes and/or obesity in 40 patients. This approach has yielded satisfactory results because for a large majority of patients, the causative mutations have been identified again. Only one patient was unable to be reconfirmed because current bioinformatics tools are limited in the detection of complex indels. Of the 39 patients identified, 3 of them are potential carriers of several causative mutations. This technique could be considered in the clinical field because it allows a multigene approach by providing a rapid diagnosis, cheaper and with a quality similar to the gold standard Sanger sequencing. For us, the purpose of this technique is a fast and optimal clinical diagnosis in order to identify unsolved cases, which are candidates for exome sequencing. This second study was published in "Diabetes Care" journal.

The third approach involves whole exome sequencing (WES) in 4 individuals where the whole family was previously tested negative for all known genes of diabetes. This approach led to the discovery of a thirteen MODY gene, *KCNJ11*, and confirms the broad phenotypic spectrum that goes from neonatal diabetes to MODY depending on the mutations. The major difficulty with this

technique is filtering variants in order to get a single causal mutation (or possibly several on the same gene) to identify new MODY genes. The strategy we used combined both a bioinformatics filter for example with filters on family cosegregation and on SNP databases and a biological filter, with the use of a technique for high-throughput genotyping. This pioneering study in the use of NGS to identify new genes of MODY has been published in "PLOS ONE".

In conclusion, this work took advantage of technological advances such as capture, targeted sequencing and NGS to elucidate and to improve the screening of monogenic forms of diabetes and obesity. This improved understanding of the molecular mechanisms may lead to the development of better treatments like personalized medicine. We hope to see direct improvements for patients in the near future, such as a more accurate, faster and more comprehensive molecular diagnostics.

# Sommaire

---

I.	Introduction.....	10
A.	Le diabète de type 2 .....	10
1)	Contexte .....	10
2)	Épidémiologie.....	10
3)	Définition.....	13
4)	Complications.....	14
5)	Critères de diagnostic du diabète.....	16
6)	Physiopathologie .....	16
7)	Traitements .....	19
8)	Facteurs de risque .....	21
9)	Causes génétiques des formes communes .....	22
10)	Causes génétiques des formes rares.....	24
B.	L'obésité .....	32
1)	Contexte .....	32
2)	Définition.....	32
3)	Épidémiologie.....	33
4)	Complications.....	35
5)	Physiopathologie .....	37
6)	Traitements .....	39
7)	Facteurs de risque .....	44
8)	Causes génétiques des formes communes .....	47
9)	Causes génétiques des formes rares.....	49
C.	Génétique des maladies rares.....	54
1)	Analyse de liaison .....	54
2)	Séquençage Sanger .....	55
3)	Séquençage nouvelle génération .....	56
II.	Présentation des travaux.....	65
A.	Article 1: Whole-Exome Sequencing and High Throughput Genotyping Identified <i>KCNJ11</i> as the Thirteenth MODY Gene.....	65
B.	Article 2: Highly sensitive diagnosis of 43 monogenic forms of diabetes or obesity through one-step PCR-based enrichment in combination with next-generation sequencing .....	69



C.	Article 3: A nonsense loss-of-function mutation in <i>PCSK1</i> contributes to dominantly inherited human obesity.....	73
III.	Discussion générale et perspective .....	78
A.	Le séquençage d'exome .....	78
B.	Le séquençage ciblé.....	80
C.	Perspectives thérapeutiques.....	81
D.	Perspectives technologiques.....	84
E.	Perspectives scientifiques .....	87
IV.	Bibliographie.....	91

## Liste des abréviations

ADA	American Diabetes Association
AGB	Adjustable Gastric Band
AMA	Association médicale américaine
CDC	Centers for Disease Control and Prevention
CNV	Copy-number variation
CVD	CardioVascular Disease
DEXA	Dual Energy X-ray Absorptiometry
DT2	Diabète de Type 2
EASD	European Association for the Study of Diabetes
EMA	European Medicines Agency
ESHG	European Society of Human Genetics
FDA	Food and Drug Administration
FID	Fédération Internationale du Diabète
GAD	Glutamate Acide Décarboxylase
GIP	Glucose-dependent Insulinotropic Polypeptide
GLP-1	Glucagon-Like Peptide 1
GWAS	Genome-wide association study
HDL	High Density Lipoprotein
HGPO	Hyperglycémie provoquée par voie orale
HLA	Human Leukocyte Antigen
IAPP	Islet Amyloid PolyPeptide
IMC	Indice de Masse Corporelle
LDL	Low Density Lipoprotein
MAF	Minor Allele Frequency
MetS	Syndrome Métabolique
MNT	Maladies Non Transmissibles
MODY	Maturity-Onset Diabetes of the Young
NAFLD	Non-alcoholic fatty liver disease
NASH	Nonalcoholic steatohepatitis
NEJM	The New England Journal of Medicine
NDM	Neonatal Diabetes Mellitus
NGS	Next generation sequencing
NHLBI	National Heart, Lung, and Blood Institute
OMS	Organisation Mondiale de la Santé
RMR	Resting Metabolic Rate
RYGB	Roux-en-Y Bypass
SNP	Single Nucleotide Polymorphism
TEE	Total Energy Expenditure
VSG	Vertical Sleeve Gastrectomy
WES	Whole-Exome Sequencing
WGS	Whole-Genome Sequencing

# I. Introduction

---

## A. Le diabète de type 2

### 1) Contexte

Bien qu'évoqué dans certains ouvrages de médecine chinoise, le diabète n'a été décrit pour la première fois que dans le papyrus Ebers datant d'environ 1500 ans av. J.-C.<sup>1</sup> et mieux défini dans le traité indien Sushruta Samhita qui daterait de 500 ans av. J.-C.<sup>2</sup>. Le terme *diabetes* vient du grec ancien et signifie *passer au travers* en référence à deux symptômes de la maladie, la soif (polydipsie) et le besoin fréquent d'uriner (polyurie). Le terme latin *mellitus* signifiant sucré comme le miel a été ajouté par le médecin anglais Thomas Willis après avoir noté le goût sucré des urines de patients diabétiques.<sup>3</sup> Si le rôle du foie dans la régulation de la glycémie a été mis en évidence en 1855 par Claude Bernard<sup>4</sup>, le rôle central du pancréas dans la pathogenèse du diabète n'a été compris que 34 ans plus tard par Mering et Minkowski<sup>5</sup>. Cette découverte constitua la base de l'isolation de l'insuline et son utilisation clinique par Banting et Best en 1921<sup>6</sup>, et fut récompensée par un prix Nobel en 1923. Depuis, dix autres prix Nobel ont été attribués en lien avec le diabète avec notamment la détermination de la séquence en acides aminés de l'insuline en 1951 par Frederick Sanger<sup>7</sup> et de la structure tridimensionnelle de l'insuline en 1969 par Dorothy Hodgkin<sup>8</sup>.

### 2) Épidémiologie

Un rapport de l'Organisation mondiale de la Santé (OMS) de 2011 stipule que les maladies non transmissibles (MNT) sont la cause la plus importante de mortalité dans le monde avec plus de 63% de la totalité des décès annuels.<sup>9</sup> En 2008, 36 millions de décès ont été imputés aux MNT (dont 1 million au diabète) et 80% d'entre eux surviennent dans les pays à revenu faible ou moyen. La fédération internationale du diabète (FID) estime que 8.3% des adultes, soit 382 millions de personnes, sont atteintes de diabète de par le monde.<sup>10</sup> La parité est presque respectée puisqu'en 2013 198 millions d'hommes vs 184 millions de femmes sont diabétiques. La FID prévoit une augmentation de 55% d'ici 2035, soit 592 millions de diabétiques. Il y a plus de diabétiques vivant en zone urbaine (246 millions) qu'en zone rurale (136 millions) et cette différence devrait s'accroître d'ici 2035. Les zones rurales des pays à revenus très faibles subiront vraisemblablement la plus forte augmentation. Certains groupes ethniques sont plus touchés par le diabète, notamment les populations indigènes. En effet, la prévalence du diabète est plus élevée chez les Amis et Atayals taïwanais, les Aborigènes d'Australie, les Sioux nord-américains et les Pimas Nord-Amérindiens.

Actuellement, la FID estime à 19,8 millions le nombre d'adultes atteints de diabète dans la région AFR (**Fig.1**), un chiffre qui devrait plus que doubler d'ici 2035. L'urbanisation grandissante et le vieillissement de la population a induit une augmentation des taux de diabète et d'obésité. Ainsi, les estimations de prévalence du diabète dans les villes rejoignent et parfois même dépassent celles des pays à haut revenu. Une grande proportion (63%) des cas de diabète sont non diagnostiqués.

De la même façon, un développement économique et social rapide couplé au vieillissement de la population a induit une forte augmentation de la prévalence du diabète de type 2 (DT2) dans la

région MENA. La FID estime à 34,6 millions le nombre de diabétiques (**Fig. 1**), chiffre qui devrait presque doubler d'ici 2035. Cette région a la plus forte prévalence de diabète avec 10,9% de diabète. D'ailleurs, trois des pays de cette région sont dans le top 10 des pays ayant la plus forte prévalence : l'Arabie Saoudite (24%), le Koweït (23%) et le Qatar (23%).

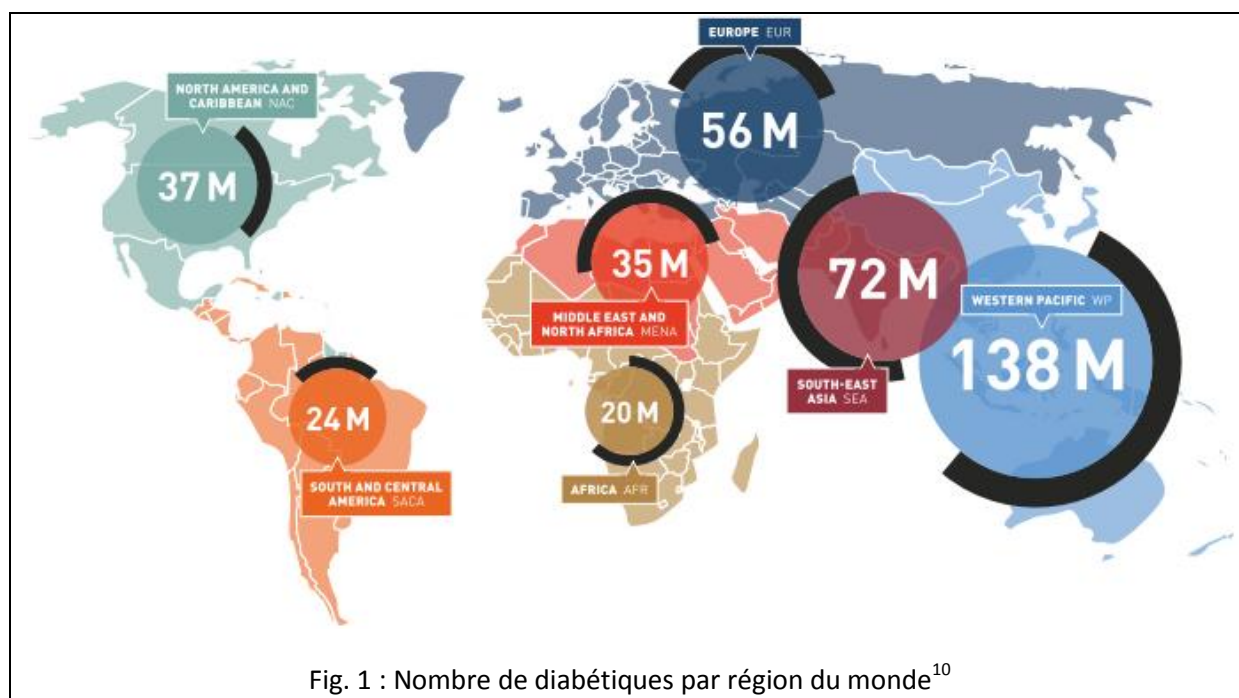
Le nombre de diabétiques dans la région EUR est estimé à 56,3 millions (**Fig. 1**), soit 8,5 % de la population adulte. La Turquie a la plus forte prévalence (14,8%). C'est la région avec la plus forte prévalence de diabète de type 1 avec environ 129300 cas et 20000 nouveaux cas par an.

La FID estime que 36,8 millions de diabétiques vivent dans la région NAC (**Fig. 1**) et que ce nombre devrait être porté à 50,4 millions en 2035. Les îles des Caraïbes sont les plus touchées et la prévalence est constamment au-dessus de la moyenne mondiale : Belize (15,9%), la Guyane (15,9%), Curaçao (14,5%) et la Martinique (14,3%). La prévalence dans la région NAC est de 9,6% ce qui en fait la deuxième région la plus touchée. Les dépenses de santé relatives au diabète sont élevées avec 42% des dépenses mondiales et dont la majeure partie provient des États-Unis et du Canada.

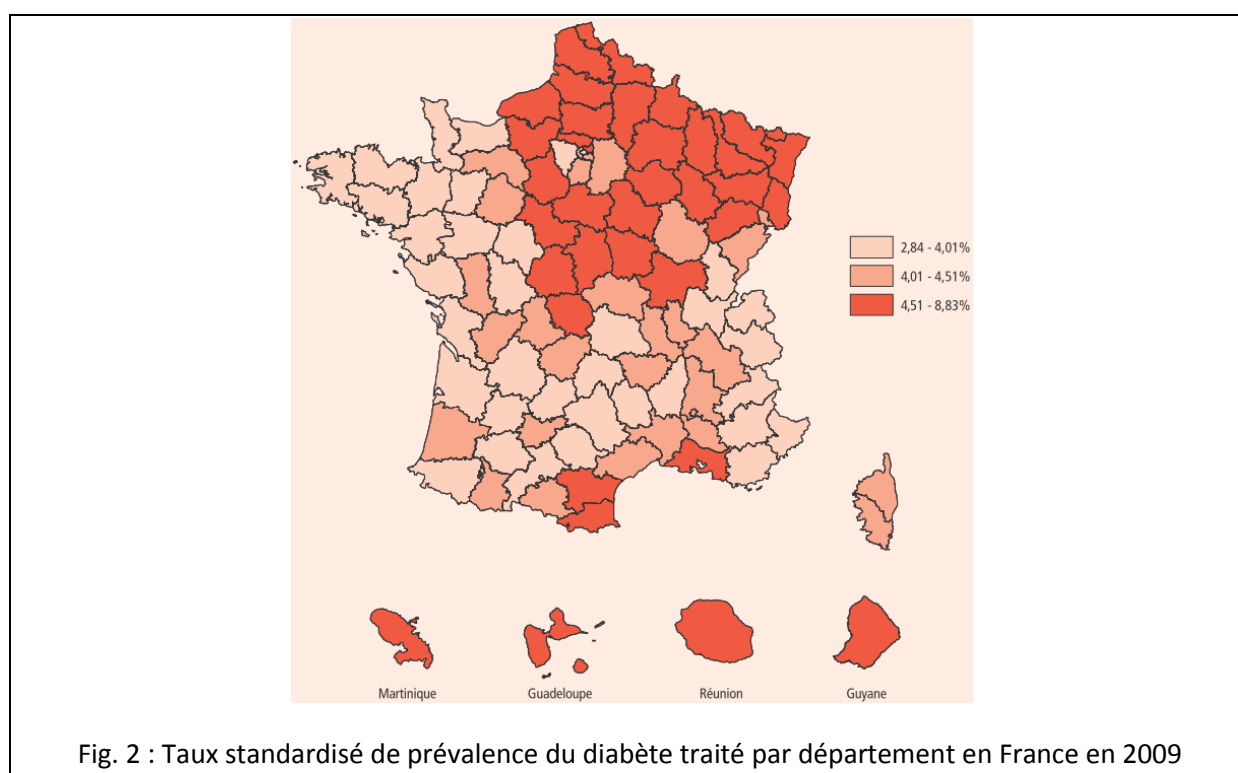
La prévalence de la région SACA est estimée à 8% de la population adulte, soit 24,1 millions de diabétiques. (**Fig. 1**) En 2035, la FID estime que ce nombre devrait augmenter de presque 60%, car cette région est en transition économique et l'urbanisation grandissante devrait influencer la prévalence.

Environ 8,2% de la population adulte mondiale, soit 72,1 millions de diabétiques vivent dans la région SEA. (**Fig. 1**) En 2035, ce nombre devrait être porté à 10,1% soit 123 millions de diabétiques. Les îles Maurice ont la plus forte prévalence de diabète (14,8%), mais aussi le plus haut PIB.

La région WP abrite 138,2 millions de diabétiques soit 8,6% des adultes. (**Fig.1**) Au cours des 20 prochaines années, l'augmentation de la prévalence est estimée à 11,1%. Les îles du Pacifique sont les plus touchées par le diabète : les Tokelau (37,5%), la Micronésie (35,0%), les îles Marshall (34,9%), les Kiribati (28,8%), et les îles Cook (25,7%).



L'étude ENTRED réalisée entre 2000 et 2009 a permis d'estimer le taux de prévalence du diabète traité en France.<sup>11</sup> Ce dernier a progressé de 2,6% à 4,4% soit de 1,6 à 2,9 millions de personnes pour la France entière. Ce taux est estimé par la FID à 8,2% pour 2035. Des variations géographiques importantes des taux de prévalence du diabète étaient observées en 2009. (**Fig. 2**) Les taux étaient particulièrement élevés dans les DOM, dans le nord et le nord-est de la métropole, notamment au Nord-Pas-de-Calais avec 5,4 % et faibles dans l'ouest de la métropole. Des taux contrastés étaient observés au sein de la région Ile-de-France avec par exemple 5,8 % en Seine-Saint-Denis, mais seulement 3,2 % à Paris. Si un risque génétique élevé permettait d'expliquer la forte prévalence dans les régions d'outre-mer, des conditions socioéconomiques défavorables sont davantage en cause pour les autres départements les plus touchés. En effet, un niveau socioéconomique défavorisé est corrélé avec une progression plus forte de l'obésité.<sup>12</sup> À noter que l'origine ethnique joue aussi un rôle puisque les femmes d'origine maghrébine avaient un risque de diabète plus de 2 fois supérieur à celui des femmes d'origine française.



La FID estime que presque la moitié (46%, soit 175 millions) des personnes ne sont pas diagnostiquées pour le diabète de par le monde. Cependant, cette prévalence varie drastiquement selon les régions du monde et serait autour de 20% en France<sup>13</sup> alors qu'elle peut atteindre 90% dans les pays à revenus très faibles. Cette proportion élevée s'explique d'une part, par un manque de moyens pour le dépistage ou un mauvais diagnostic dû à la grande variété de complications, et d'autre part, par le fait que les stades de prédiabète peuvent durer plusieurs années en restant asymptomatiques. De plus, la difficulté d'appliquer des résultats de recherche du traitement du DT2 et des changements de comportement de quelques patients à une population demeure un véritable défi.

Les frais de santé relatifs au diabète représentent 10,8% des dépenses totales liées à la santé dans le monde en 2013, soit environ 548 milliards de dollars. La menace de submerger les systèmes

de santé à travers le monde est bien réelle sachant que ce coût pourrait excéder 627 milliards de dollars d'ici 2035. Ces dépenses sont loin d'être homogènes et 20% des dépenses globales sont faites dans les pays à revenus faibles et moyens, là où 80% des diabétiques résident. En moyenne, 5621\$/an sont dépensés par diabétique dans les pays à hauts revenus alors que 356\$/an sont dépensés dans les pays à revenus moyens et faibles. En France, le remboursement moyen annuel était estimé en 2007 à 5300€ par diabétique. En 2007, l'hôpital et les médicaments représentaient les plus gros postes de dépenses, soit 37 % (4,7 milliards €) et 27% (3,4 milliards €) des dépenses totales.<sup>14</sup>

### 3) Définition

Le diabète est défini par l'Association Américaine du Diabète (ADA) comme un groupe de maladies métaboliques caractérisées par une hyperglycémie résultant de défauts de sécrétion à l'insuline, d'action de l'insuline (insulino-résistance), ou des deux.<sup>15</sup> L'ADA scinde le diabète en 4 grandes catégories.

Dans la première catégorie, on trouve le diabète de type 1 qui compose 5 à 10% des cas de diabète. Précédemment appelé diabète insulino-dépendant, il résulte d'une destruction auto-immune des cellules  $\beta$  du pancréas. Les marqueurs biologiques de cette destruction, détectés dans 85 à 90% des cas, peuvent être des autoanticorps dirigés contre les cellules  $\beta$ , contre l'insuline, contre le glutamate acide décarboxylase (GAD) et les tyrosines phosphatases IA-2 a et IA-2 b. L'origine de ce désordre auto-immun est liée en partie à la prédisposition génétique (les allèles HLA-DR/DQ) et à des facteurs environnementaux qui sont encore peu connus. Dans quelques rares cas, ce diabète peut être dû à des mutations délétères dans les gènes *AIRE*<sup>16,17</sup>, *FOXP3*<sup>18,19</sup>, *SIRT1*<sup>20</sup> ou *STAT3*<sup>21</sup>. Il existe également une forme rare appelée diabète idiopathique, car on ne connaît pas l'origine de la destruction des cellules  $\beta$ .

Dans la deuxième catégorie, on trouve le diabète de type 2 qui compose la majeure partie des cas de diabète, soit 90 à 95%. On le nommait avant diabète non insulino-dépendant ou diabète de l'adulte, car cette forme de diabète ne nécessite pas de traitement à l'insuline dans la plupart des cas et est diagnostiquée en général autour de 45 ans bien que cet âge moyen ait tendance à diminuer au cours des années. La cause de ce diabète n'est pas liée à une destruction des cellules  $\beta$ , mais plutôt à une résistance à l'insuline et/ou de défaut de sécrétion d'insuline.

La troisième catégorie définie par l'ADA est un peu fourre-tout, car on y trouve 8 sous-types de diabète. Dans la première sous-catégorie, on trouve des formes rares qui composent environ 1 à 2% des cas de DT2 et qu'on appelle Maturity-Onset Diabetes of the Young (MODY) et diabète néonatal. Contrairement au DT2 commun polygénique, ces formes de diabète sont dues à des altérations génétiques fortes, le plus souvent d'un seul gène, qui conduisent à un dysfonctionnement de la cellule  $\beta$  et un défaut de sécrétion d'insuline. Cette classification de l'ADA est discutable dans le sens où même si ces formes sont rares et débutent précocement, elles répondent pleinement à la définition du DT2.

Dans les autres sous-catégories, on trouve d'autres formes particulières de diabète qui peuvent être dues à :

- Des formes d'insulino-résistance de type A: le syndrome de Rabson-Mendenhall (dû à des mutations délétères dans le gène INSR<sup>22,23</sup>), le lépréchaunisme, les lipodystrophies, etc.
- Des maladies pancréatiques exocrines : pancréatite, néoplasie, mucoviscidose, etc.
- Des endocrinopathies : hyperthyroïdie, glucagonome, syndrome de Cushing, etc.
- Des médicaments ou des produits chimiques : Vacor (contre les rongeurs), pentamidine (antiparasitaire), diazoxide (vasodilatateur), etc.
- Des infections virales : rubéole congénitale, cytomégalovirus, etc.
- Des formes d'insulino-résistance de type B : syndrome de l'homme raide, etc.
- Des maladies syndromiques : syndrome de Wolfram, de Prader-Willi, Bardet-Biedl, etc.

Dans la dernière catégorie, on trouve le diabète gestationnel. Pendant longtemps, le diabète gestationnel était défini comme un prédiabète diagnostiqué pendant la grossesse. La définition s'appliquait même si le prédiabète se résolvait avec l'accouchement ou était présent, mais non diagnostiqué avant le début de la grossesse. Ces limitations impliquaient un biais important dans la classification du diabète gestationnel d'autant plus que le nombre de femmes enceintes non diagnostiquées, mais atteintes de DT2 n'a cessé de croître ces dernières années. En effet, l'ADA estime qu'environ 7% des grossesses (entre 1 et 14% selon la population étudiée) s'accompagnent d'un diabète gestationnel soit environ 200 000 cas par an. Le groupe international IADPSG a été mis en place et un consensus a été trouvé en 2009.<sup>24</sup> En 2011, l'ADA recommande que les femmes à haut risque dont le diabète est diagnostiqué à la première visite prénatale (6-12 semaines) par une HGPO (Hyperglycémie provoquée par voie orale) rentrent dans la catégorie diabète de type 2 et non du diabète gestationnel.<sup>25</sup> Sur la base d'un rapport récent du National Institutes of Health (NIH)<sup>26</sup>, l'ADA a modifié ses recommandations pour le diagnostic du diabète gestationnel en 2014. Sur la base du consensus IADPSG, on procède à une HGPO à 24-28 semaines chez des femmes non diabétiques. Sinon le processus en deux étapes recommandé par le NIH peut être utilisé : à 24-28 semaines, on fait une épreuve de glucose sans être à jeun avec 50g de glucose. Si la glycémie à 1h est supérieure à 140 mg/dL alors une HGPO est réalisée pour établir fermement le diagnostic de diabète gestationnel. À l'heure actuelle, on ne sait pas si l'une des deux stratégies est meilleure que l'autre tant en termes de sensibilité que de rapport coût/bénéfice. Le risque de diabète consécutif à un diabète gestationnel est élevé<sup>27</sup> et l'ADA recommande un dépistage tous les 3 ans.<sup>28</sup> Le dépistage reste important dans le sens où une grande partie des cas est résolue avec une modification du style de vie ou de la prise de metformine.<sup>28</sup> Cependant, on ne connaît pas les réels bénéfices d'un tel traitement sur la mère, par exemple une possible diminution du risque de diabète et une amélioration des paramètres métaboliques. On sait toutefois que le taux de macrosomie fœtale est diminué.<sup>29,30</sup>

#### 4) Complications

Les symptômes de l'hyperglycémie marquée incluent une polyurie (le besoin fréquent d'uriner), une polydipsie (la soif), une perte de poids et parfois une polyphagie et des troubles de la vue.<sup>15</sup> Un retard de croissance et une susceptibilité à certaines infections peuvent aussi accompagner l'hyperglycémie chronique.<sup>15</sup> Les conséquences aiguës d'un diabète non contrôlé sont une

hyperglycémie avec acidocétose ou un syndrome d'hyperosmolarité hyperglycémique qui peut mener à un coma.<sup>15</sup>

Les complications à long terme du diabète incluent<sup>15</sup> :

- Une rétinopathie avec une potentielle perte de la vue
- Une néphropathie pouvant mener à une insuffisance rénale
- Une neuropathie périphérique avec un risque d'ulcère des pieds, d'amputation ou au pied de Charcot (ostéo-arthropathie nerveuse)
- Une neuropathie autonome causant des symptômes gastro-intestinaux, génito-urinaires (avec parfois une dysfonction sexuelle), et cardiovasculaires

Le diabète est la première cause de cécité acquise, d'amputation non traumatique et d'insuffisance rénale.<sup>31</sup> Cependant, une étude anglaise récente a montré que le diabète n'était plus que la deuxième cause de cécité, derrière les maladies héréditaires de la rétine, peut-être grâce à l'intensification des programmes de dépistage.<sup>32</sup> Le diabète est aussi la deuxième cause de maladies cardiovasculaires.<sup>31</sup> Le risque est effectivement doublé pour cette cause et il s'explique par une athérosclérose qui peut aussi conduire à un accident vasculaire cérébral ou une artériopathie oblitérante des membres inférieurs.<sup>33</sup> Par ailleurs, le diabète s'accompagne fréquemment de problèmes d'hypertension et d'anomalies du métabolisme des lipoprotéines. En effet, la prévalence de l'hypertension est estimée à plus de 50% des patients atteints de DT2.<sup>34</sup>

Une autre cause de morbidité et de mortalité est le risque augmenté de cancer chez les diabétiques. Plusieurs études épidémiologiques montrent que les diabétiques ont une incidence augmentée pour plusieurs types de cancers : foie, voies biliaires, pancréas, estomac, colon, reins, vessie, sein et endomètre. (**Fig. 3**) De façon opposée, l'incidence est réduite pour le cancer de la prostate. Les mécanismes mis en jeu dans ces associations ne sont pas élucidés, mais certains éléments sont mis en cause comme l'hyperinsulinémie, la résistance à l'insuline, une dérégulation de l'hormone de croissance, l'obésité, le traitement médical ou encore le régime suivi. L'étude deCODE réalisée sur 44655 individus a montré que le hazard ratio de la mortalité due au cancer était de 1,12 chez les prédiabétiques, 1,28 chez les diabétiques récemment diagnostiqués et 1,57 chez ceux avec un diabète connu.<sup>35</sup> Une autre méta-analyse a montré un risque de 1,27 chez les diabétiques après ajustement pour l'IMC.<sup>36</sup>

Type de cancer	Risque relatif
Foie	2,01
Voies biliaires	1,43
Pancréas	1,94
Estomac	1,18 (chez les femmes)
Colon	1,27
Reins	1,42
Vessie	1,24
Sein	1,2
Endomètre	2,1
Prostate	0,84

Fig. 3 : Cancers associés au diabète<sup>37</sup>



## 5) Critères de diagnostic du diabète

En 1997, l'ADA révisa les critères de diagnostic du diabète sur la base d'études épidémiologiques transversales qui mettaient en relation rétinopathie et glycémie à jeun/HGPO/HbA1c (hémoglobine glyquée).<sup>38</sup> (Fig. 4) Si le seuil de 2 g/l fut conservé pour une hyperglycémie provoquée par voie orale (HGPO), le seuil de la glycémie à jeun est passé de 1,4 à 1,26 g/l. L'HbA1c est utilisée comme marqueur de la glycémie reflétant le niveau de glucose des 2-3 derniers mois. L'HbA1c résulte de la glycation de la chaîne  $\beta$  des molécules d'hémoglobine transportant le dioxygène au sein des globules rouges. L'HbA1c est devenue l'indice le plus utilisé pour évaluer, chez les patients diabétiques, l'efficacité des traitements antidiabétiques sur le contrôle glycémique à long terme.<sup>39</sup> Ce marqueur, auparavant utile pour la gestion du patient, peut maintenant être aussi utilisé pour le diagnostic du diabète. Cependant, pour des raisons de standardisation du test, mais aussi de coût, il commence seulement à être utilisé en Europe et depuis 2009 aux USA. Ce test comporte des avantages tels que la commodité d'usage (pas besoin d'être à jeun) et la robustesse (peu de variabilité due au stress et/ou la maladie). Cependant, ce test est basé sur un facteur sanguin, l'hémoglobine, et est donc soumis à un biais dans le cas d'hémoglobinopathie ou d'anémie par exemple. Dans ce cas, seuls la glycémie à jeun et l'HGPO peuvent être utilisés. Une étude a montré qu'à l'échelle de la population générale et en prenant le seuil d'HbA1c  $\geq 6,5$  % pour diagnostiquer un DT2, environ 2 % de la population auraient un faux diagnostic de DT2.<sup>40</sup>

	HbA1c (en%)	Glycémie à jeun (en g/l)	HGPO (en g/l)
Diabète	$\geq 6,5$	$\geq 1,26$	$\geq 2$
Prédiabète	5,7 à 6,4	1 à 1,25	1,4 à 1,99
Normal	$\leq 5,7$	$\leq 0,99$	$\leq 1,39$

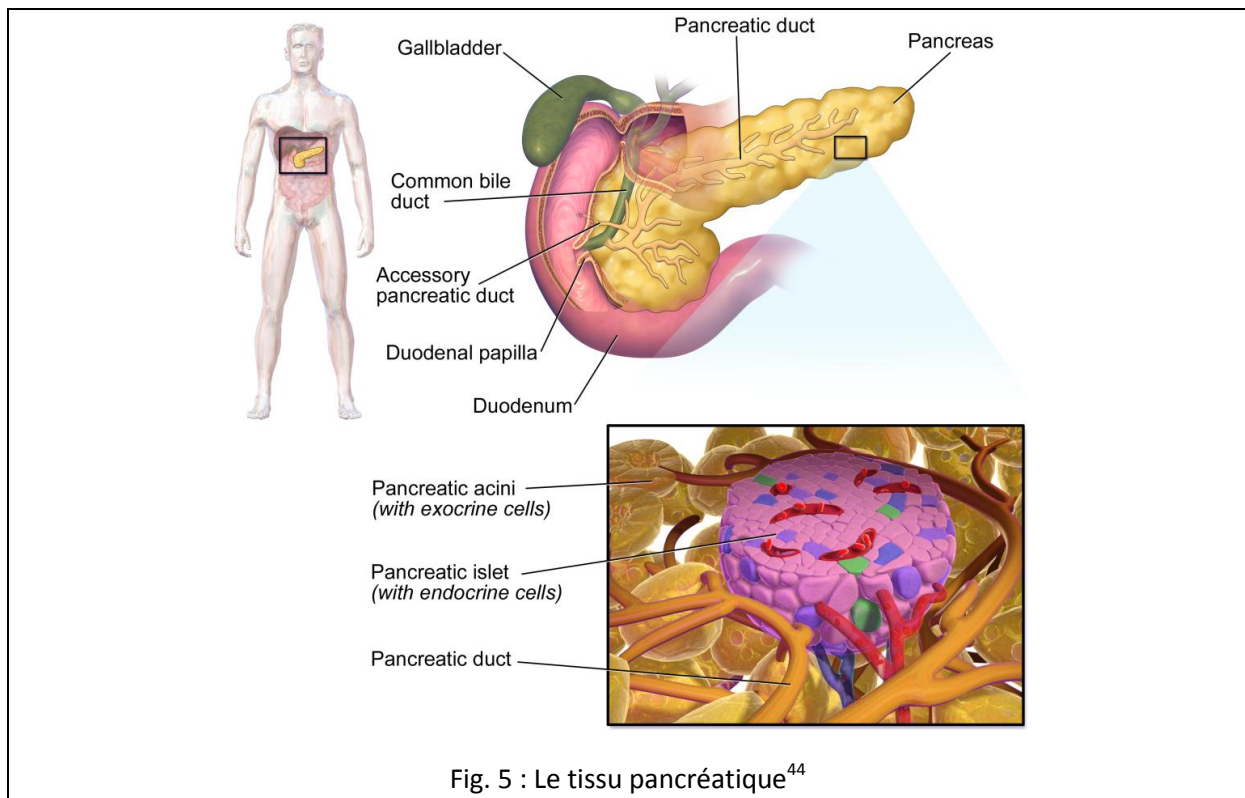
Fig. 4 : Critères de l'ADA pour le diagnostic du diabète

## 6) Physiopathologie

### a) Physiologie du pancréas

Le foie, les ovaires, les testicules mais aussi le pancréas sont des glandes dites amphicrines (ou mixtes), car ces glandes possèdent à la fois des fonctions exocrines et endocrines. Les glandes exocrines délivrent leur sécrétion par l'intermédiaire d'un canal excréteur vers le milieu extérieur de l'organisme alors que les glandes endocrines libèrent directement leurs sécrétions dans la circulation sanguine au niveau des capillaires sanguins. La partie exocrine du pancréas est composée de cellules acineuses qui synthétisent les enzymes digestives (amylase, trypsine, lipase, etc.) et de cellules canalaire ou ductales qui produisent une fraction importante de la sécrétion hydroélectrolytique, riche en bicarbonate de sodium, pour réguler le pH du suc pancréatique.<sup>41</sup> (Fig.5) Ce suc pancréatique liquide est déversé ensuite vers le duodénum par un mécanisme complexe faisant intervenir de nombreux transporteurs et canaux ioniques.

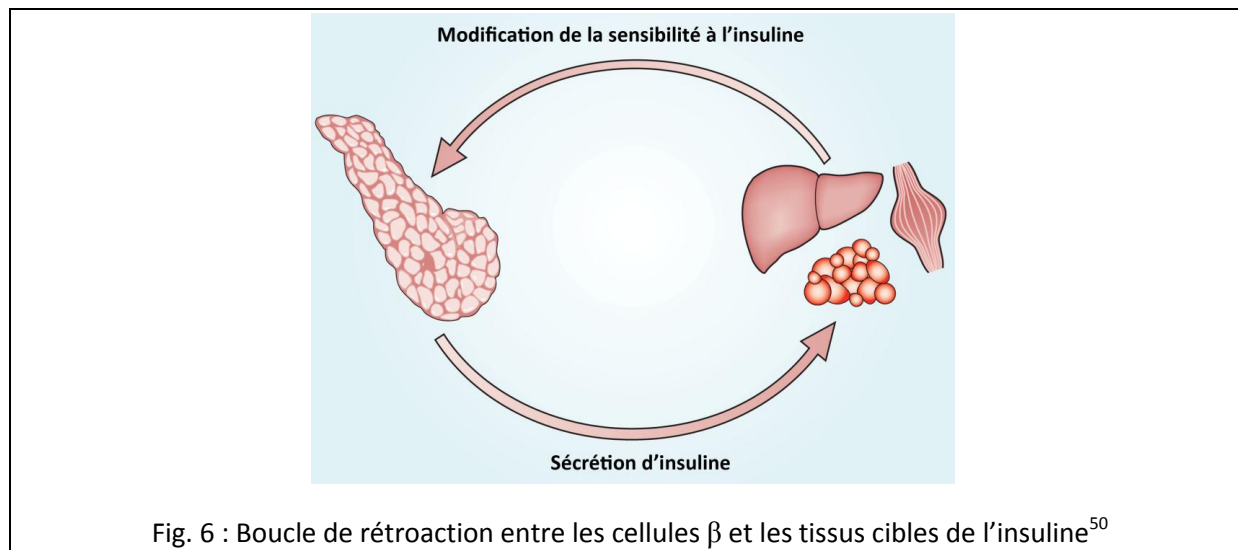
Regroupées en amas appelés îlots de Langerhans et représentant environ 1% de sa masse, les cellules endocrines du pancréas sont constituées principalement de deux types de cellules :  $\alpha$  et  $\beta$ . L'architecture cellulaire, soit la quantité de cellules et leurs localisations, varie entre espèces. Par exemple, chez la souris les cellules  $\beta$  sont regroupées dans le centre de l'îlot et les cellules  $\alpha$  à la périphérie alors que chez l'homme les cellules  $\alpha$  et  $\beta$  sont dispersées le long des vaisseaux sanguins en association avec la microcirculation.<sup>42</sup> Les cellules  $\beta$  représentent environ 75% de la masse endocrine du pancréas et produisent l'insuline et l'amyline ou IAPP (islet amyloid polypeptide), la première étant sécrétée en réponse à une élévation du taux de glucose plasmatique.<sup>43</sup> Les cellules  $\alpha$ , qui constituent environ 20% de la masse endocrine du pancréas, sécrètent le glucagon en réponse à une baisse de la glycémie.<sup>43</sup>



#### b) Physiopathologie du pancréas

L'importance de la résistance à l'insuline et du dysfonctionnement de la cellule  $\beta$  a longtemps été débattue. On pensait au départ que la résistance à l'insuline était l'anomalie principale du DT2 et que l'incapacité à sécréter l'insuline était une manifestation tardive.<sup>45</sup> Cette notion changea avec la découverte que dans la plupart des systèmes endocrines humains, une boucle de rétroaction permet d'assurer l'homéostasie du glucose en limitant les variations.<sup>46</sup> Cette boucle de rétroaction repose sur un dialogue entre les cellules  $\beta$  et les tissus cibles de l'insuline. (**Fig. 6**) L'insuline libérée en réponse à la stimulation des cellules  $\beta$  sert d'intermédiaire à la consommation du glucose, des acides aminés et des acides gras par les tissus cibles de l'insuline. En retour, ces tissus envoient l'information aux îlots de Langerhans de leur besoin d'insuline. L'acteur majeur de ce processus n'a à ce jour pas été identifié, mais certaines pistes indiquent un rôle possible de la betatrophine<sup>47</sup>, du cerveau via la leptine<sup>48</sup>, ou même du système immunitaire<sup>49</sup>. Dans le cas d'une résistance à l'insuline,

comme dans une bonne partie des obèses spécialement si l'adiposité est viscérale, la cellule  $\beta$  augmente la production d'insuline pour maintenir un taux de glucose constant. Cependant, si la cellule  $\beta$  ne peut accomplir cette tâche, la glycémie augmente et évolue progressivement d'une intolérance au glucose au diabète de type 2. Cette progression est clairement hétérogène et varie selon de nombreux facteurs pathogènes.



Plusieurs mécanismes permettraient d'expliquer ce phénomène, par exemple la réduction de masse  $\beta$ -pancréatique.<sup>51</sup> Les raisons de cette perte sont multifactorielles et incluent la glucolipotoxicité<sup>52</sup> et le dépôt des plaques amyloïdes qui résultent de l'apoptose des cellules  $\beta$  par une augmentation du stress oxydatif au niveau du réticulum endoplasmique<sup>53</sup>. Cette perte n'est pas compensée par le remplacement de nouvelles cellules,  $\beta$  car le pancréas semble incapable de se régénérer après 30 ans.<sup>54</sup> Cependant, aucune étude ne quantifie le lien entre la masse de cellules  $\beta$  et leur fonctionnalité. En effet, un défaut de fonctionnalité de la cellule  $\beta$  pourrait amorcer cette apoptose et des études supplémentaires seront nécessaires pour approfondir ce mécanisme.

Bien que moins étudiée, la dérégulation de la sécrétion de glucagon par les cellules  $\alpha$  contribue au développement de l'hyperglycémie.<sup>55</sup> Cette dérégulation se manifeste sous forme d'une augmentation de la concentration de glucagon à jeun et d'une incapacité à diminuer la sécrétion de glucagon après l'ingestion d'un repas. Il reste à définir si cette dérégulation est un changement primaire au niveau de la cellule  $\alpha$  ou secondaire suite à une anomalie de la cellule  $\beta$ . Cependant, le flux sanguin va des cellules  $\beta$  vers les cellules  $\alpha$  puis vers les cellules  $\delta$  produisant la somatostatine ; de fortes concentrations en insuline sont capables d'inhiber la sécrétion de glucagon dans les cellules  $\alpha$ .<sup>56</sup> D'autres produits de la cellule  $\beta$ , tels que le zinc, l'acide  $\gamma$ -aminobutyrique (GABA) et le glutamate, pourraient aussi réguler la sécrétion de glucagon.<sup>56</sup>

Le tractus gastro-intestinal produit plusieurs peptides dont certains régulent directement l'absorption des nutriments : le Glucagon-like peptide 1 (GLP-1) et le peptide insulinothéropé dépendant du glucose (GIP), connus sous le nom d'incrétines. Le GLP-1 a un rôle plus important et agit à la fois sur les cellules  $\beta$  pour augmenter la sécrétion d'insuline et sur les cellules  $\alpha$  pour inhiber la sécrétion de glucagon.<sup>57</sup> Les concentrations plasmatiques de GLP-1 étant similaires chez des individus normoglycémiques et DT2, on suppose que c'est la réponse des cellules  $\beta$  au GLP-1 qui doit être défaillante.<sup>58</sup> Ce résultat est en phase avec le modèle de déficience globale de la cellule  $\beta$  en

réponse à d'autres molécules telles que les sulfonylurées ou les agonistes  $\beta$ -adrénergiques.<sup>59</sup> D'autre part, après une chirurgie bariatrique chez des DT2, on observe une augmentation de GLP-1 plasmatique qui est bénéfique pour le patient.<sup>60</sup> Cependant, d'autres mécanismes pourraient aussi expliquer cette amélioration de la glycémie.<sup>61</sup>

Le microbiome intestinal semble avoir aussi un rôle important dans la pathophysiologie du DT2.<sup>62</sup> Le métagénome humain comporte environ 100 fois plus d'informations qu'un seul génome humain. Le microbiome joue un rôle important dans plusieurs maladies, dont le DT2, mais on ne connaît pas précisément la ou les espèces bactériennes qui causent un changement de métabolisme.<sup>63</sup> Deux études basées sur l'utilisation d'échantillons fécaux suggèrent que des changements fonctionnels du microbiome intestinal pourraient être directement liés au développement du DT2.<sup>64,65</sup> Cependant, les marqueurs métagénomiques diffèrent entre Européens et Chinois suggérant que la capacité de prédire le développement du DT2 varie selon la population.<sup>65</sup> Une autre étude a montré des améliorations de la sensibilité à l'insuline chez des patients atteints de syndrome métabolique 6 semaines après une greffe du microbiome intestinal de sujets normopondéraux.<sup>66</sup> La flore intestinale agirait sur l'absorption des nutriments et s'adapterait rapidement à un changement de régime alimentaire.<sup>67</sup> Enfin, le microbiome intestinal a permis d'expliquer une intolérance au glucose chez des souris lors de la prise d'édulcorants de synthèse.<sup>68</sup> L'hypothèse est que certaines bactéries de l'intestin réagissent aux édulcorants en sécrétant des substances qui provoquent une réponse inflammatoire comme lors d'une overdose de sucre, ce qui développe des changements dans la capacité à utiliser le sucre.

Le système nerveux est un autre régulateur important des processus métaboliques. Le système nerveux parasympathique contrôle directement le métabolisme du glucose via le nerf vague et le système sympathique indirectement via l'hypothalamus. Ce dernier a un rôle central dans le sens où il agit sur la sécrétion de glucagon et d'insuline<sup>67</sup> ainsi que sur la production de glucose hépatique<sup>69</sup> par la détection du glucose<sup>69</sup>, de l'insuline<sup>70</sup> et des acides gras circulants<sup>71</sup>. D'autres éléments du cerveau sont importants dans le maintien des rythmes circadiens et du sommeil.<sup>72</sup> On peut citer le gène *CLOCK*<sup>73</sup> qui régule l'horloge biologique et la mélatonine<sup>74</sup> qui est d'ailleurs appelée hormone du sommeil.

## 7) Traitements

Depuis la découverte de l'insuline, les approches médicales pour la prévention et le traitement du diabète ont été transformés par les avancées scientifiques qui ont significativement amélioré la vie des diabétiques et allongé leur durée de vie. Ces avancées résultent d'études cliniques comme l'utilisation de l'insuline humaine biosynthétique, des seringues et aiguilles petites et faciles d'usage, le suivi de la glycémie<sup>75</sup>, et les pompes à insuline<sup>76</sup>. La prévention du diabète passe par une prise en charge à tous les niveaux : médecins, infirmiers, nutritionnistes, travailleurs sociaux, podologues, etc. Certains programmes de prévention utilisant cette approche multidisciplinaire ont montré que l'activité physique et la perte de poids peuvent réduire le risque de diabète de 58% chez des personnes prédisposées.<sup>77</sup> Des effets majeurs peuvent être aussi apportés par le traitement à la metformine et aux sulfonylurées. L'insulinothérapie intensive reste le traitement de référence pour prévenir l'hyperglycémie chez les patients les plus atteints.<sup>78</sup> Au Royaume-Uni, la proportion de patients traités à l'insuline a été multipliée par 3 entre 1991 et 2010.<sup>79</sup> De plus dans ce même pays, la

proportion de patients prenant l'insuline seule est passée de 97% dans la première décennie à 37% dans la seconde.<sup>79</sup> Ce simple fait implique un rôle grandissant des agents oraux utilisés en combinaison avec l'insuline.

Il existe 3 autres classes d'antidiabétiques très utilisés : les thiazolidinediones (toutefois retirées du marché en France), et les analogues du GLP-1 et les inhibiteurs de la DPP-4 qui sont des incrétines. Ces traitements peuvent être utilisés en combinaison jusqu'à 3 classes différentes. Les recommandations de l'ADA et de l'EASD (European Association for the Study of Diabetes) sont résumées dans la **figure 7**.<sup>80</sup> Concrètement, on commence chez la plupart des patients par un changement de mode de vie couplé à un traitement à la metformine. Si un taux correct d'HbA1c n'est pas atteint après trois mois ou si l'HbA1c est initialement supérieure à 9.0%, on passe à une bithérapie, c'est-à-dire que la metformine est couplée à un autre antidiabétique. Suivant la stratégie utilisée, on peut optimiser le traitement par rapport au patient par exemple en limitant les hypoglycémies, la prise de poids ou les effets secondaires ou tout simplement en minimisant les coûts. Si le taux d'HbA1c n'est toujours pas correct après 3 mois, alors une trithérapie est envisagée et l'insuline reste le traitement le plus efficace surtout quand l'HbA1c est élevée.

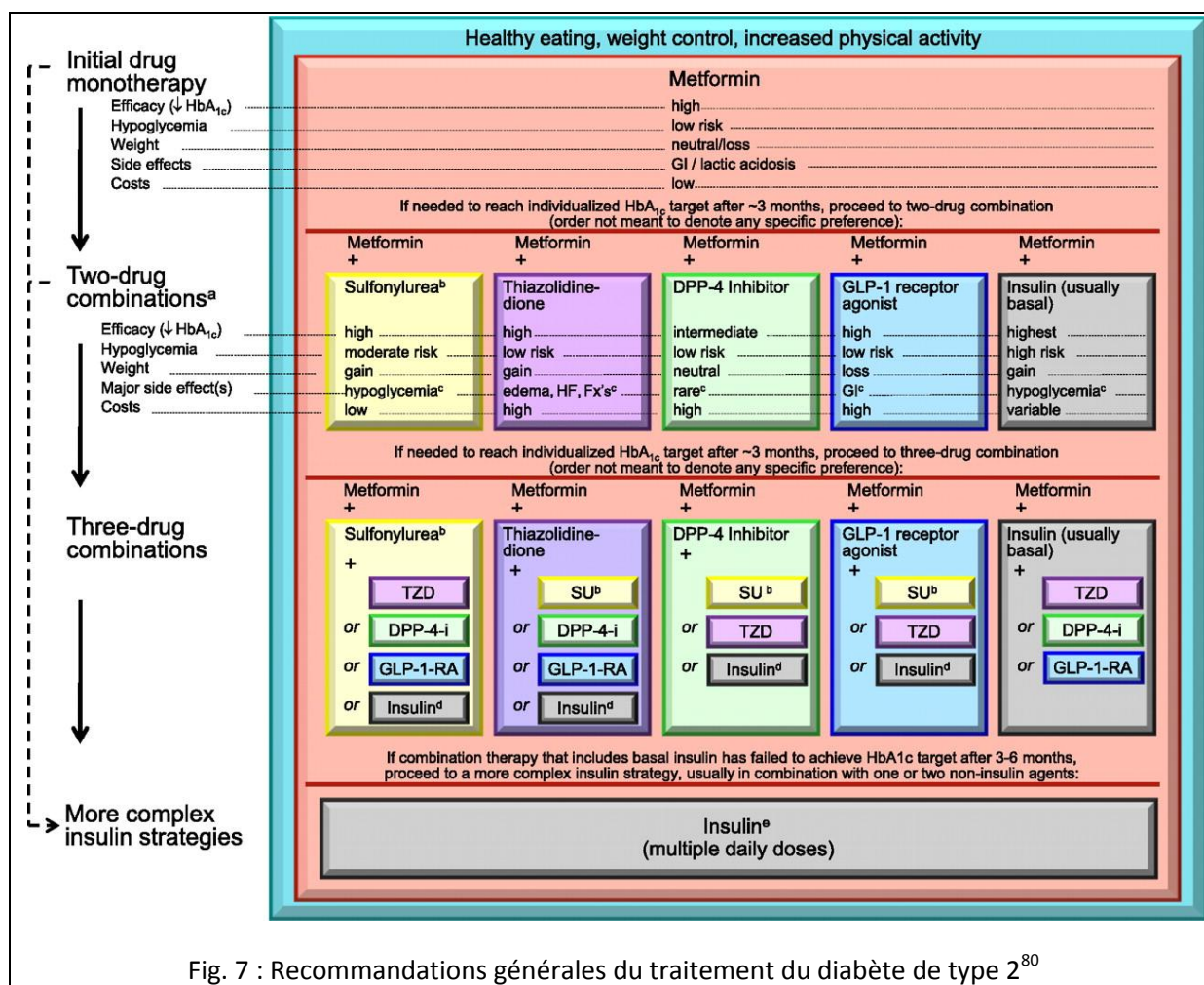


Fig. 7 : Recommandations générales du traitement du diabète de type 2<sup>80</sup>

Les stratégies préventives et les traitements des complications ne sont pas en reste avec plusieurs progrès à la clé concernant :



- néphropathie : antagonistes des récepteurs de l'angiotensine, inhibiteurs de l'Enzyme de Conversion de l'Angiotensine,<sup>81</sup> et simplement la restriction protéique<sup>82</sup>
- insuffisance rénale : greffe de reins
- rétinopathie : photocoagulation au laser<sup>83</sup>

Des avancées récentes de la greffe de cellules  $\beta$  et de pancréas ont récemment démontré une guérison chez une majorité des patients traités.<sup>84</sup> La chirurgie bariatrique dans un contexte d'obésité permet une perte de poids qui améliore les paramètres sanguins tels que la glycémie et qui va parfois jusqu'à une rémission du DT2.<sup>85,86</sup>

De nouvelles approches sont développées et plusieurs traitements sont encore en phase expérimentale en exploitant différentes stratégies<sup>50</sup> :

- augmenter la fonction sécrétoire de la cellule  $\beta$
- augmenter la masse de cellules  $\beta$
- diminuer/bloquer l'effet du glucagon
- réduire la production hépatique de glucose
- augmenter l'action de l'insuline
- combiner deux peptides pour obtenir 2 effets, par exemple une double incrétine

## 8) Facteurs de risque

Beaucoup d'études ont établi des associations entre plusieurs facteurs de risque et le DT2. L'IMC<sup>87</sup>, des niveaux anormaux de lipides<sup>88</sup>, l'hypertension<sup>89</sup>, le tabagisme<sup>90</sup>, l'alimentation<sup>91</sup>, l'inactivité physique<sup>92</sup>, la sédentarité<sup>93</sup>, les antécédents familiaux de diabète<sup>94</sup>, un prédiabète ou un diabète gestationnel<sup>27</sup>, l'inflammation<sup>95</sup>, l'environnement intra-utérin<sup>96</sup>, l'âge<sup>94</sup>, le sexe<sup>97</sup>, l'ethnicité<sup>87</sup> et le statut socioéconomique<sup>98</sup> sont les plus fréquemment documentés.

L'épidémie globale de DT2 est en partie liée aux taux de surpoids et d'obésité parmi les adultes, mais aussi chez les enfants. En effet, plus de 50% des patients atteints de diabète de type 2 sont obèses et plus de 30% en surpoids.<sup>99</sup> Le surpoids et l'obésité sont les meilleurs prédicteurs du DT2 et cet effet est plus important chez les jeunes adultes. En effet, la prise de poids chez le « jeune » adulte (25-40 ans) implique un risque de DT2 plus grand et l'âge de début semble précoce par rapport à une prise de poids après 40 ans.<sup>100</sup> Une des hypothèses est que le dépôt de graisse dans le foie et dans le pancréas favorise le développement d'une insulino-résistance hépatique et le dysfonctionnement de la cellule  $\beta$ .<sup>101,102</sup>

Une hypothèse récente implique le développement intra-utérin. C'est une période critique durant laquelle le milieu intra-utérin peut affecter négativement le développement fœtal par des modifications épigénétiques.<sup>103</sup> Ces modifications impliquent notamment la méthylation de l'ADN et des histones sans changer la séquence d'ADN, mais son expression. Ainsi, un faible poids de naissance est associé à un risque augmenté de développer un DT2.<sup>104</sup> Une des hypothèses est que ce faible poids de naissance est dû à une malnutrition/sous-nutrition qui implique le développement d'un phénotype « économe » où le métabolisme du glucose est altéré.<sup>105</sup> Ce phénomène serait accentué dans le cas où l'environnement inverse (surabondance alimentaire) serait présent durant le développement de l'enfant impliquant un gain de poids rapide.<sup>106,107</sup> Le milieu intra-utérin peut aussi

être modifié lors d'un diabète gestationnel. Les femmes avec un diabète gestationnel ont un risque 7 fois élevé de développer un DT2 comparé à des femmes qui n'ont pas d'hyperglycémie durant leur grossesse.<sup>108</sup> De plus, ces expositions fœtales à une hyperglycémie maternelle prédisposent l'enfant à un développement précoce d'un DT2.<sup>109</sup> Le mécanisme biologique pourrait passer par une modification de la méthylation de certains gènes et augmenterait la susceptibilité de développer un DT2<sup>110</sup>, mais aussi d'autres maladies complexes telles que l'obésité<sup>111</sup>, des maladies métaboliques<sup>112</sup>, des complications cardiovasculaires<sup>113</sup> et même du cancer<sup>114</sup>.

Enfin, d'autres facteurs ont été associés indépendamment avec le risque de DT2. On peut citer le taux de vitamine D<sup>115</sup>, la dépression<sup>116</sup> ou l'utilisation d'antidépresseurs<sup>117</sup>, le stress au travail<sup>118</sup>, et les troubles du sommeil. En lien avec les troubles du sommeil, le syndrome d'apnées obstructives du sommeil est lié au développement du diabète.<sup>119</sup> Suivant la sévérité du syndrome, les patients ont entre 23 et 30% d'augmentation du risque de développer un diabète qui pourrait être lié à la désaturation en oxygène, la privation de sommeil (diminution du temps de sommeil paradoxal) et au rythme cardiaque moyen plus élevé. Egalement lié aux troubles du sommeil, un travail posté augmente de 9% le risque de développer un diabète et de 37% uniquement chez les hommes.<sup>120</sup>

D'autres études suggèrent un rôle potentiel des toxines environnementales comme les perturbateurs endocriniens et des particules dans l'air. Parmi ces contaminants environnementaux, on peut citer les polychlorobiphényles (PCBs), les composés organochlorés (OC), les pesticides, le bisphénol A, les phtalates et tout dernièrement l'acide perfluorooctanoïque (PFOA).<sup>121</sup>

## 9) Causes génétiques des formes communes

Si des facteurs environnementaux sont mis en cause dans le développement du diabète, la part de génétique héritable joue également un rôle. L'héritabilité est définie comme la proportion de variation phénotypique dans une population qui est attribuable à la variation génétique entre les individus. Une façon indirecte de montrer qu'un caractère est héritable est de calculer le risque de développer une maladie et de comparer les risques de patients dont les parents sont atteints par rapport au risque dans une population générale. Ainsi, un enfant a 35% de chance de développer un diabète si un parent est atteint (et davantage si c'est la mère plutôt que le père) et 70% si les deux parents sont atteints en comparaison d'environ 10% dans la population générale.<sup>122,123</sup> Cependant, ce risque est déterminé par des facteurs de risque génétiques et environnementaux ce qui ne permet pas de séparer ces deux composantes. Le caractère héritable du diabète et d'autres maladies peut être estimé de façon assez précise par les études de jumeaux monozygotes et dizygotes. L'hypothèse principale est que les jumeaux partagent le même environnement prénatal et postnatal, et que la part génétique similaire est de 100% pour les jumeaux monozygotes et de 50% pour les jumeaux dizygotes. Les estimations d'héritabilité du DT2 sont d'environ 70% chez les jumeaux monozygotes et de 20-30% chez les jumeaux dizygotes.<sup>124,125</sup> Cette mesure de l'héritabilité est variable selon les études et s'expliquent par différent biais comme l'âge<sup>126</sup> ou encore le caractère génétique pas si unique chez les jumeaux monozygotes<sup>110,127</sup>. La contribution exacte de la part génétique reste à déterminer dans le DT2. En effet, alors que les allèles HLA expliquent jusqu'à 50% des cas d'agrégation familiale du diabète de type 1<sup>128</sup>, aucun locus majeur de susceptibilité au diabète de type 2 n'a été trouvé.

Depuis 2003, le séquençage du génome humain et le projet international HapMap ont permis de générer des cartes fines du génome humain divisé en des milliers de groupes de polymorphismes nucléotidiques (SNPs) fréquents hautement corrélés, c'est-à-dire montrant un fort déséquilibre de liaison (lorsque deux mutations sont toujours associées sur une même séquence). L'objectif principal du projet HapMap était de comparer les séquences génétiques de différents individus, issus de diverses populations afin de relever les régions chromosomiques où des variations génétiques sont partagées. La phase II du projet avait démontré que la variabilité des SNPs détectés avec une fréquence de l'allèle mineur  $\geq 5\%$  pouvait être résumée en  $\sim 550\,000$  blocs de déséquilibre de liaison chez les individus originaires d'Europe ou d'Asie. Chaque bloc de LD est représenté par un tagSNP.

Simultanément, une véritable révolution technologique a eu lieu dans le génotypage des SNPs. En moins de 10 ans, les avancées technologiques ont permis de passer du test d'un seul SNP au génotypage simultané de centaines de milliers de SNPs par individu, en utilisant des puces à ADN. Par les études d'association pangénomique ou *genome-wide association studies* en anglais (GWAS), les généticiens ont mis ces puces à ADN à profit afin d'identifier des différences de fréquence allélique des tagSNPs présents sur les puces entre des cas, c'est-à-dire des patients atteints de DT2, versus des contrôles normoglycémiques. (**Fig. 8**) Une différence de fréquence significative entre allèles du même SNP entre cas et témoins indique que la région correspondante du génome contient des variants génomiques fonctionnels qui influencent le risque de DT2. Une autre stratégie des GWAS consiste à étudier la variation génétique de traits quantitatifs reliés à la maladie étudiée. (**Fig. 8**)

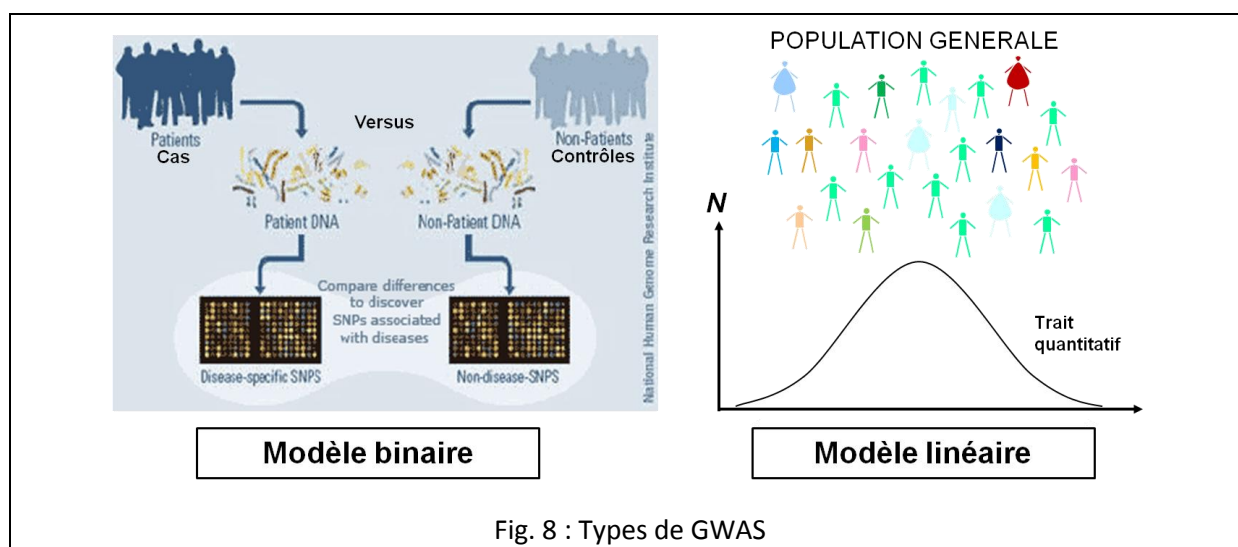


Fig. 8 : Types de GWAS

Ces études d'association pangénomique ont identifié à ce jour 90 variants génétiques qui augmentent le risque de DT2.<sup>129</sup> Le locus de *TCF7L2* contient le variant ayant l'effet le plus fort (odds ratio : 1,46).<sup>130</sup> Alors que la majorité de ces loci ont été découverts lors de méta-analyses du consortium DIAGRAM, les loci de traits quantitatifs relatifs au diabète ont été davantage étudiés par le consortium MAGIC. Aujourd'hui, des traits tels que la glycémie à jeun, l'insulinémie à jeun, le taux de glucose à 2h après une HGPO ou encore l'HbA1c ont permis la mise en évidence de 72 loci associés.<sup>129</sup> Certains de ces loci sont aussi associés au DT2, mais pas tous. Ces chevauchements partiels posent la question des mécanismes sous-jacents conduisant au DT2. Ce résultat peut par exemple être dû à la difficulté de quantifier la sécrétion d'insuline, un manque de puissance statistique (étant donné l'effet faible de certains SNPs), mais aussi à cause de la multiplicité des mécanismes de pathogenèse du DT2. Des études fonctionnelles ont déjà démontré des mécanismes



liés au glucose, au GLP-1, à l'exocytose de l'insuline, à la biogenèse des granules d'insuline ou encore des mécanismes post-transcriptionnels de maturation de l'insuline.<sup>129</sup>

Même si les GWAS ont été réalisés sur des populations d'origine européenne, de nouvelles études sur d'autres ethnies commencent à émerger. Par exemple, deux nouveaux loci, *KCNQ1* et *C2CD4A* ont été associés au DT2 chez les japonais<sup>131-133</sup> et un certain nombre d'autres loci chez des Asiatiques de l'est<sup>134,135</sup>. De façon générale, les loci associés au DT2 sont relativement conservés entre les différentes ethnies.<sup>136,137</sup> Une étude utilisant 4 groupes ethniques différents a montré que les effets des SNPs communs n'atteignant pas le seuil de significativité sont homogènes entre les groupes et cette méta-analyse a révélé 7 nouveaux loci.<sup>138</sup> Enfin, notons qu'une étude a conduit avec succès chez des Groenlandais un GWAS qui a mis en évidence une mutation non-sens dans le gène *TBC1D4* qui explique 15% des cas de diabète dans ce pays.<sup>139</sup> L'augmentation du risque est très élevée puisque chez des cas homozygotes, 60% des plus de 40 ans et 80% des plus de 60 ans ont développé un DT2.

Cependant, ces variants expliquent seulement 10% de l'héritabilité de la maladie et augmentent le risque de diabète de 10 à 30%.<sup>140</sup> La multiplicité des loci qui contribuent au risque de DT2 fait que l'on qualifie le DT2 de polygénique. Cette complexité rend difficile le calcul du risque et le développement de stratégies préventives et thérapeutiques basé sur le profil génétique. Cependant, les approches méthodologiques ne se limitent pas aux GWAS, et l'ère du séquençage nouvelle génération devrait apporter des réponses à la part d'héritabilité manquante du DT2. En effet, les GWAS ne testent qu'une hypothèse principale : la forte incidence du DT2 est due aux SNPs fréquents. Dès lors, on écarte comme possibles causes les variants rares, les zones du génome non couvertes par des SNPs fréquents ainsi que les mécanismes épigénétiques. De plus, le chromosome X, Y et l'ADN mitochondrial avaient été écartés des analyses et de nouvelles études commencent seulement à en tenir compte.<sup>141</sup>

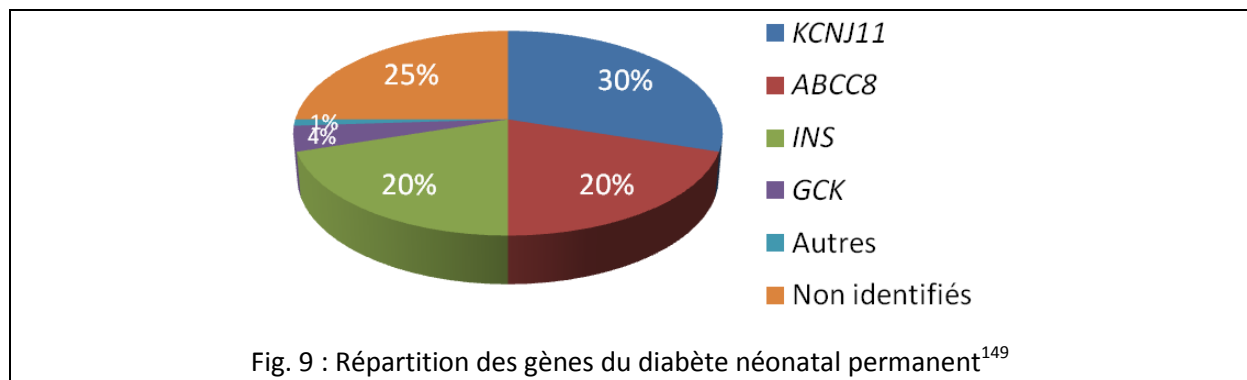
## 10) Causes génétiques des formes rares

### a) Le diabète néonatal

Le diabète néonatal (NDM) est défini comme un diabète seul ou accompagné d'autres symptômes, diagnostiqué durant les 6 premiers mois de la vie. Cette forme de diabète est très rare et touche environ 1/100 000 à 1/250 000 des naissances.<sup>142</sup> Ce diabète peut être permanent (PNDM) ou transitoire (TNDM) dans le cas où il y a rémission, généralement dans les 18 premiers mois de vie. Dans la plupart des cas de diabète néonatal, les parents ne sont pas diabétiques et cela s'explique par un événement génétique *de novo*, c'est-à-dire uniquement présent chez l'enfant. Cependant dans certains cas, les parents ne sont pas atteints, mais porteurs du même défaut génétique que l'enfant. On connaît à l'heure actuelle une vingtaine de gènes qui peuvent être la cause de diabète néonatal. (**Fig. 10**)

Les causes les plus fréquentes de PNDM sont des mutations dans les gènes qui codent pour le canal potassique, *ABCC8* et *KCNJ11*, et le gène de la préproinsuline, *INS*. (**Fig.9**) Le gène *GCK* qui code pour la glucokinase et le gène *PDX1* qui code pour un facteur de transcription sont les 2 seuls autres gènes connus qui causent uniquement un diabète néonatal. Les autres cas identifiés de

diabète néonatal sont des formes très rares qui sont dites syndromiques, car elles ne touchent pas qu'au pancréas et sont la cause de phénotypes complexes avec de nombreuses anomalies : cardiaques, neurologiques, rénales, intestinales, développementales, etc. Les découvertes les plus récentes soulignent la présence de mutations dans des gènes codant pour des facteurs de transcription. Par exemple, les gènes *EIF2AK3*<sup>143</sup>, *GATA4*<sup>144</sup>, *GATA6*<sup>145</sup>, *PDX1*<sup>146</sup>, *PTF1A*<sup>147</sup> et *RFX6*<sup>148</sup> sont mutés chez des patients atteints d'aplasie, d'hypoplasie ou encore d'agénésie du pancréas ce qui souligne un rôle essentiel de ces gènes pour le bon développement et la différenciation du pancréas.



Les mutations présentes dans les gènes *ABCC8* et *KCNJ11* qui conduisent au diabète néonatal sont dites activatrices, car elles conduisent à une suractivité du canal potassique. Cependant, les mutations sont parfois délétères et conduisent à une perte de fonction du canal potassique.<sup>150</sup> Cette maladie s'appelle hyperinsulinisme congénital isolé (CHI) et l'hyperinsulinémie implique une hypoglycémie constante qui parfois ne peut être corrigée que par traitement chirurgical. D'autres gènes peuvent être touchés dans cette maladie, par exemple si les gènes *GCK* ou *GLUD1* contiennent des mutations gain de fonction.<sup>151,152</sup> Citons également les gènes *HADH*, *HNF4A*, *SLC16A1* et *UCP2* qui conduisent au CHI.<sup>153-156</sup>

Le TNDM représente 50 à 60% des cas de diabète néonatal.<sup>157</sup> L'hyperglycémie transitoire s'accompagne généralement d'un retard de croissance ce qui souligne le rôle central de l'insuline dans le développement du jeune enfant et la nécessité d'un traitement à l'insuline.<sup>157</sup> Chez une partie des patients, un stade de prédiabète ou même de diabète peut subsister jusqu'à l'âge adulte.<sup>157</sup> La plupart des cas de TNDM sont dus à une anomalie du chromosome 6q24. Il s'agit d'une surexpression d'un allèle qui peut être due à 3 mécanismes différents<sup>158</sup> :

- Une isodisomie uniparentale de la région 6q24 : les 2 allèles présents chez l'enfant proviennent du père
- Une duplication de la région 6q24 héritée du père
- Hypométhylation de la région 6q24 : l'allèle maternel n'est plus méthylé ce qui altère son expression

Cette région d'environ 400kb contient 2 gènes : *HYMA1* qui code pour un ARNm non traduit et *ZAC* qui régule *PACAP1*, codant pour un récepteur présent dans le pancréas qui jouerait un rôle dans la sécrétion d'insuline.<sup>158</sup>

Environ 5% des cas de TNDM sont dus à des mutations récessives dans le gène *ZFP57*, causant une hypométhylation dans plusieurs loci.<sup>159</sup> Le phénotype induit est complexe et inclut en

plus du TNDM un retard de croissance intra-utérin, une macroglossie, des malformations cardiaques et un retard de développement.<sup>160</sup> Le phénotype est très variable, car certains individus seront sévèrement atteints et d'autres ne seront pas affectés.<sup>160</sup>

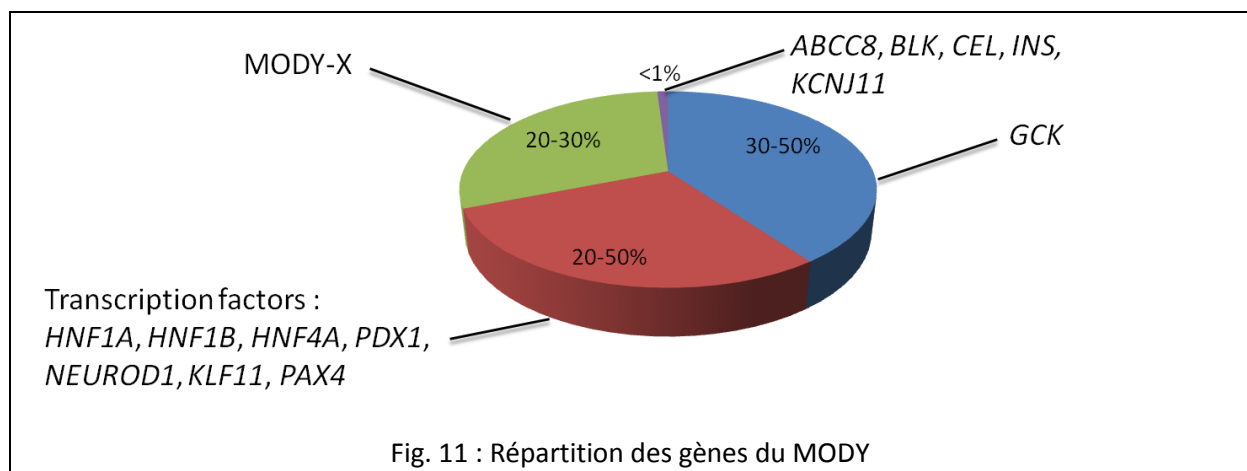
Gène	Protéine (UniProt)	Modèle	Phénotype	Caractéristiques cliniques associées	Principes
<i>ABCC8</i>	ATP-binding cassette sub-family C member 8	dominant	TNDM, PNDM	syndrome DEND ou iDEND, parfois RCIU	150
<i>EIF2AK3</i>	Eukaryotic translation initiation factor 2-alpha kinase 3	récessif	PNDM	syndrome de Wolcott-Rallison	161
<i>GATA4</i>	Transcription factor GATA-4	récessif	PNDM	RCIU, malformations cardiaques, retard de développement	144
<i>GATA6</i>	Transcription factor GATA-6	dominant	PNDM, TNDM	malformations cardiaques congénitales	145
<i>GCK</i>	Glucokinase	récessif	PNDM	RCIU	162
<i>GLIS3</i>	Zinc finger protein GLIS3	récessif	PNDM	hypothyroïdie, choléstase, glaucome, polykystose rénale, infections récurrentes, surdité	163
<i>IER3IP1</i>	Immediate early response 3-interacting protein 1	récessif	PNDM	Microcéphalie, encéphalopathie épileptique	164
<i>INS</i>	Insulin	dominant	PNDM	retard de croissance intra-utérin (RCIU)	165
<i>KCNJ11</i>	ATP-sensitive inward rectifier potassium channel 11	dominant	PNDM, TNDM	syndrome DEND ou iDEND	166
<i>MNX1</i>	Motor neuron and pancreas homeobox protein 1	récessif	PNDM	RCIU sévère	167
<i>NEUROD1</i>	Neurogenic differentiation factor 1	récessif	PNDM	diarrhée intractable	168
<i>NEUROG3</i>	Neurogenin-3	récessif	PNDM	RCIU, diarrhée	169
<i>NKX2-2</i>	Homeobox protein Nkx-2.2	récessif	PNDM	RCIU, retard de développement sévère	170
<i>PAX6</i>	Paired box protein Pax-6	récessif	PNDM	anomalies des yeux et du cerveau	171
<i>PCBD1</i>	Pterin-4-alpha-carbinolamine dehydratase	récessif	PNDM	hyperphénylalaninémie	172
<i>PDX1</i>	Pancreas/duodenum homeobox protein 1	récessif	PNDM	insuffisance pancréatique exocrine	146
<i>PTF1A</i>	Pancreas transcription factor 1 subunit alpha	récessif	PNDM	hypoplasie du cervelet	147
<i>RFX6</i>	DNA-binding protein RFX6	récessif	PNDM	atrésie intestinale, aplasie de la vésicule biliaire, diarrhée intractable	148
<i>SLC19A2</i>	Thiamine transporter 1	récessif	PNDM ou DM plus tardif	Syndrome TRMA, anémie mégaloblastique, surdité neurosensorielle	173
<i>SLC2A2</i>	Solute carrier family 2, facilitated glucose transporter member 2	récessif	TNDM, PNDM	syndrome de Fanconi-Bickel, dysfonctionnement rénal, hypergalactosémie	174
<i>WFS1</i>	Wolframin	récessif	DM autour de 5/6 ans	syndrome de Wolfram, atrophie optique, surdité	175

Fig. 10 : Causes génétiques de NDM

DEND : developmental delay, epilepsy, and neonatal diabetes, iDEND : intermediate DEND, RCIU : retard de croissance intra-utérin, TRMA : thiamine-responsive megaloblastic anemia

## b) Le diabète MODY

Le terme MODY (Maturity Onset Diabetes of the Young) est né en 1975 lorsque Tattersall décrivit une forme modérée de diabète avec transmission dominante dans 3 familles.<sup>176</sup> En effet, les patients qu'il avait observés pouvaient se passer d'insuline et étaient relativement jeunes par rapport à des patients atteints de DT2. En étudiant d'autres familles, le mode de transmission apparaissait comme autosomique. On sait maintenant que ce type de diabète se développe chez l'enfant, l'adolescent ou le jeune adulte. Les études génétiques ont permis de mettre en évidence différents sous-types de MODY qui ont chacun des caractéristiques métaboliques et cliniques particulières. Ce type de diabète est rarement associé à l'obésité et les phénotypes au sein d'une même famille peuvent être extrêmement disparates.



En 1992, deux groupes, celui de Philippe Froguel et celui d'Andrew Hattersley, ont montré de façon indépendante par une étude de liaison qu'un gène du chromosome 7 était probablement en cause dans ce type de diabète.<sup>177,178</sup> La même année, des mutations furent identifiées dans le gène de la glucokinase qui a un rôle clé dans la régulation du glucose.<sup>162</sup> En 1996, ces deux mêmes équipes ont identifié un autre gène codant pour un facteur de transcription et étant responsable du MODY3.<sup>179</sup> La même année, l'équipe de Graeme Bell identifia une mutation non-sens dans le gène *HNF4A*, codant pour un autre facteur de transcription.<sup>180</sup> Ces deux facteurs de transcription jouent un rôle dans la régulation des cellules du foie, mais ils sont également exprimés dans les îlots pancréatiques et d'autres tissus. En 1997, deux autres facteurs de transcription, *PDX1* et *HNF1B* permettent d'expliquer les formes de MODY 4 et 5.<sup>181,182</sup> Aujourd'hui, 13 sous-types de MODY sont connus. (Fig. 12) Les causes les plus fréquentes sont des mutations dans les gènes *GSK* et *HNF1A* et de façon moins répandues dans les gènes *HNF4A* et *HNF1B*. Des mutations dans les 9 autres gènes sont très rares et environ 30% des cas restent non élucidés. (Fig. 11)

Des études récentes ont montré que dans une famille japonaise, le gène est associé au gène *EEA1*, mais les patients ne répondent pas strictement à la définition du MODY, car le diabète est déclaré plus tardivement (~37 ans).<sup>183</sup> Un poster publié à l'European Society of Human Genetics (ESHG) en 2013 par une équipe italienne revendique le gène *MAFA* comme nouveau gène du MODY. Cependant, ce gène est connu pour causer le diabète chez la souris depuis 2005<sup>184</sup> et cette étude n'est pas encore publiée. De plus, une étude a montré que des patients mutés dans ce même gène n'étaient pas atteints de diabète.<sup>185</sup> Une étude belge récente a montré une nouvelle forme de diabète syndromique liée au gène *TRMT10A*.<sup>186</sup> Il s'agit d'une famille consanguine où les 3 enfants

sont homozygotes et atteints d'un diabète qui s'accompagne d'une petite taille et une microcéphalie avec une déficience intellectuelle. Les parents qui sont hétérozygotes ont développé tous les deux un diabète à 58 ans sans d'autres symptômes. Enfin, un poster présent à l'EASD 2014 relate une nouvelle forme de DT2 familiale liée au gène *APPL1*, dont les mutations coségrègent dans 2 familles différentes et semblent fonctionnelles dans les lignées cellulaires testées.<sup>187</sup> Ce gène avait déjà été séquencé dans d'autres études de MODY-X et s'était révélé négatif.<sup>188</sup>

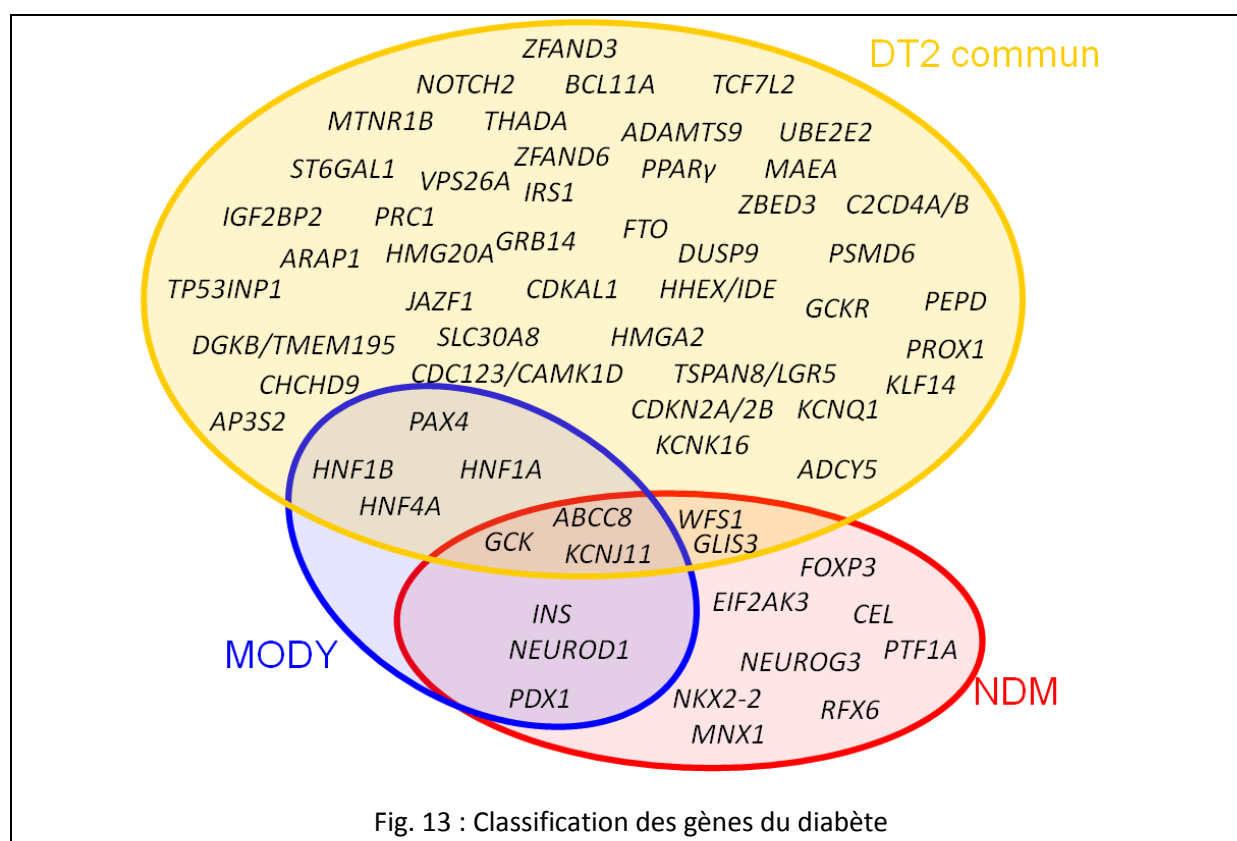
Il existe une dernière forme qui est un peu différente du MODY dans le sens où la transmission ne peut-être que maternelle, car elle est liée à des gènes qui touchent la mitochondrie. Cette forme de diabète se développe chez le jeune adulte et est systématiquement accompagnée de surdit  neurosensorielle. Il s'agit du syndrome de Ballinger-Wallace qui est d  dans la majorit  des cas   une mutation ponctuelle dans le g ne mitochondrial *MT-TL1*, qui code l'ARN mitochondrial de transfert de la leucine, et dans de rares cas   des mutations ponctuelles des g nes mitochondriaux *MT-TE* et *MT-TK* qui codent respectivement l'ARN mitochondrial de transfert de l'acide glutamique et de la lysine ou encore du g ne *MT-ND6*.<sup>189-192</sup>

G�ne	Prot�ine (UniProt)	Ph�notype	Caract�ristiques cliniques associ�es	Princeps
<i>HNF4A</i>	Hepatocyte nuclear factor 4-alpha	MODY1	complications microvasculaires, r�ductions de la concentration s�rique des triglyc�rides, des apolipoprot�ines CIII et AII, et la Lp(a)	180
<i>GCK</i>	Glucokinase	MODY2	Hyperglyc�mie mod�r�e	162
<i>HNF1A</i>	Hepatocyte nuclear factor 1-alpha	MODY3	complications microvasculaires, glycosurie r�nale, augmentation de la sensibilit� aux sulfonyleur�es, augmentation du rapport proinsuline/insuline dans le s�rum	179
<i>PDX1</i>	Pancreas/duodenum homeobox protein 1	MODY4		181
<i>HNF1B</i>	Hepatocyte nuclear factor 1-beta	MODY5	kystes r�naux et d'autres anomalies du d�veloppement r�nal, insuffisance r�nale, anomalies g�nitales (chez la femme)	182
<i>NEUROD1</i>	Neurogenic differentiation factor 1	MODY6	st�atorrh�e et douleurs abdominales intermittentes	168
<i>KLF11</i>	Krueppel-like factor 11	MODY7		193
<i>CEL</i>	Bile salt-activated lipase	MODY8	douleurs abdominales et insuffisance pancr�atique exocrine	194
<i>PAX4</i>	Paired box protein Pax-4	MODY9		195
<i>INS</i>	Insulin	MODY10		165
<i>BLK</i>	Tyrosine-protein kinase Blk	MODY11	ob�sit�	196
<i>ABCC8</i>	ATP-binding cassette sub-family C member 8	MODY12	Bonne r�ponse aux sulfonyleur�es	150
<i>KCNJ11</i>	ATP-sensitive inward rectifier potassium channel 11	MODY13	Bonne r�ponse aux sulfonyleur�es	197

Fig. 12 : Sous-types de MODY

### c) Lien entre diabète néonatal, MODY et DT2 commun

Bien que classifiés différemment, le diabète néonatal, MODY et le DT2 ont des points en communs. En effet, il est intéressant de voir que les gènes *ABCC8*, *GCK*, *KCNJ11* sont responsables d'un large spectre clinique, car ils sont associés au DT2, mais ils sont également la cause de diabète néonatal et MODY.<sup>198,199</sup> Il est également important de noter que la plupart des gènes du NDM et du MODY ne sont pas restreints à un seul groupe, mais plutôt partagés entre les différentes étiologies. **(Fig. 13)** Cette composante génétique commune ne permet pas d'expliquer toutes les formes de diabète, mais indique que des bases biologiques similaires sont partagées entre les différentes formes. Cette composante est peut-être sous-estimée ou surestimée dans le sens où on référence souvent le gène le plus proche comme étant la cause pour les SNPs associés lors des GWAS. Cependant, une partie des variants agit de façon lointaine en cis et se situent dans des séquences activatrices appelées amplificateurs (ou enhancers) où des facteurs de transcriptions se fixent. Ces séquences forment des domaines de chromatine tridimensionnelle qui peuvent être perturbés par des variants associés au diabète de type 2.<sup>200</sup> Il est donc difficile de savoir quel gène est impacté par un variant précis sans faire d'études fonctionnelles approfondies.

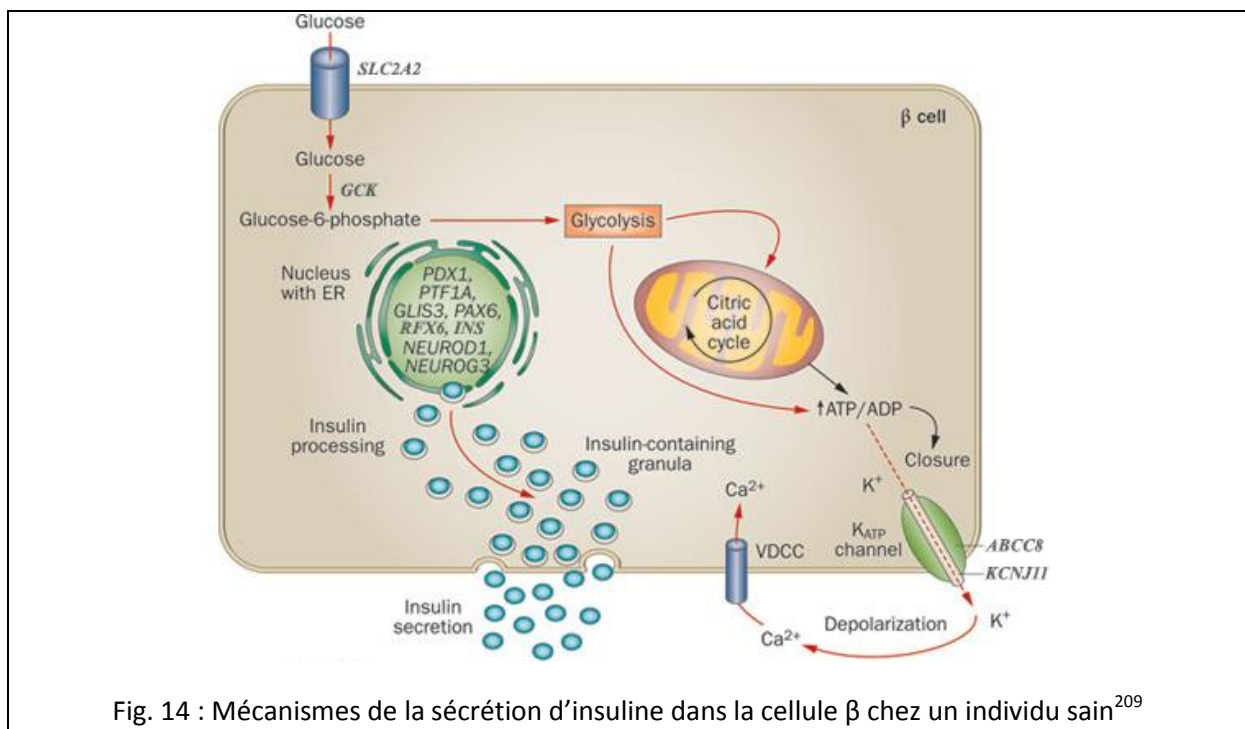


Même si une composante commune existe entre ces différentes étiologies de diabète, les phénotypes observés sont différents et les mécanismes pour arriver à la maladie diffèrent peut-être aussi. On peut par exemple hypothétiser qu'une lente détérioration de la fonction métabolique augmentant avec l'âge conduit à un stade prédiabétique où une légère hypoglycémie produit une dérégulation de gènes critiques pour conduire progressivement au déclin de la cellule  $\beta$ .<sup>201</sup> Une étude a tenté de classer les SNPs associés au DT2 en étudiant les indices de maturation de la proinsuline, de la sécrétion d'insuline et de la sensibilité à l'insuline.<sup>202</sup> Il en ressort 5 groupes dont le dernier qui n'est associé à aucun trait glycémique. De façon surprenante, ce groupe contient des gènes connus

des formes monogéniques de diabète dont *HNF1A*, *HNF1B*, *WFS1*, et *KCNJ11*. Ce manque d'association avec les traits glycémiques peut être dû à une puissance statistique trop faible ou au fait que les SNPs associés au DT2 ne jouent un rôle que dans le cas où le diabète commence à se développer.<sup>203</sup> Ces résultats soulèvent d'importantes questions mécanistiques qui ne pourront être résolues que par des études ultérieures.

#### d) Rôle biologique des gènes identifiés dans les formes rares de DT2

Chez un individu sain, le glucose est transporté dans la cellule  $\beta$  par GLUT-2 (*SLC2A2*). (**Fig. 14**) La glucokinase (*GCK*) phosphoryle ensuite le glucose en glucose-6-phosphate.<sup>204</sup> Dans le foie, cette enzyme affecte la capacité de stocker le glucose sous forme de glycogène.<sup>204</sup> Dans la cellule  $\beta$ , la glucokinase est limitante dans le métabolisme du glucose, car elle conditionne la glycolyse.<sup>204</sup> La glycolyse par le biais du cycle de Krebs permet notamment d'augmenter le ratio ATP/ADP au niveau de la membrane mitochondriale.<sup>205</sup> Cette énergie produite est détectée par les canaux potassiques et provoque leur fermeture.<sup>206</sup> La dépolarisation qui s'ensuit provoque l'ouverture des canaux calciques ; cet afflux de calcium déclenche l'exocytose de l'insuline.<sup>207</sup> Les gènes impliqués dans le transport du glucose, la glycolyse, le métabolisme mitochondrial et la sécrétion de l'insuline sont régulés par des facteurs de transcription qui conditionnent le bon développement des tissus embryonnaires, mais aussi adultes.<sup>208</sup>



Chez un individu présentant une forme monogénique de diabète, différents défauts génétiques de la cellule  $\beta$  peuvent avoir un effet direct ou indirect sur la sécrétion d'insuline. (**Fig. 15**) Une mutation délétère dans *SLC2A2* conduit à une diminution de l'afflux de glucose dans la cellule  $\beta$ .<sup>210</sup> Si *GCK* est déficient, le défaut de phosphorylation du glucose réduit la production d'ATP et réduit le stockage du glycogène dans le foie.<sup>211,212</sup> Des mutations activatrices du canal potassique (*ABCC8*, *KCNJ11*) provoquent une hyperpolarisation de la membrane bloquant l'exocytose de



l'insuline.<sup>213,214</sup> Des mutations dans le gène *INS* conduisent à une proinsuline anormale qui ne peut pas être maturée normalement dans le réticulum endoplasmique.<sup>215</sup> Cette accumulation provoque une toxicité de la cellule  $\beta$  et *in fine* son apoptose. Des défauts dans *HNF4A*, *EIF2AK3* et *WFS1* augmentent aussi l'apoptose de la cellule  $\beta$  via d'autres mécanismes.<sup>216-218</sup> Des mutations dans les gènes codant pour les facteurs de transcription *HNF1B*, *PDX1*, *PTF1A*, *RFX6*, *GLIS3*, *PAX6*, *NEUROD1* et *NEUROG3* causent un développement anormal du pancréas ou des cellules endocrines.<sup>219-226</sup> Des altérations de deux autres facteurs de transcription, *HNF1A* et *HNF4A*, diminuent la sensibilité au glucose et son métabolisme.<sup>227,228</sup>

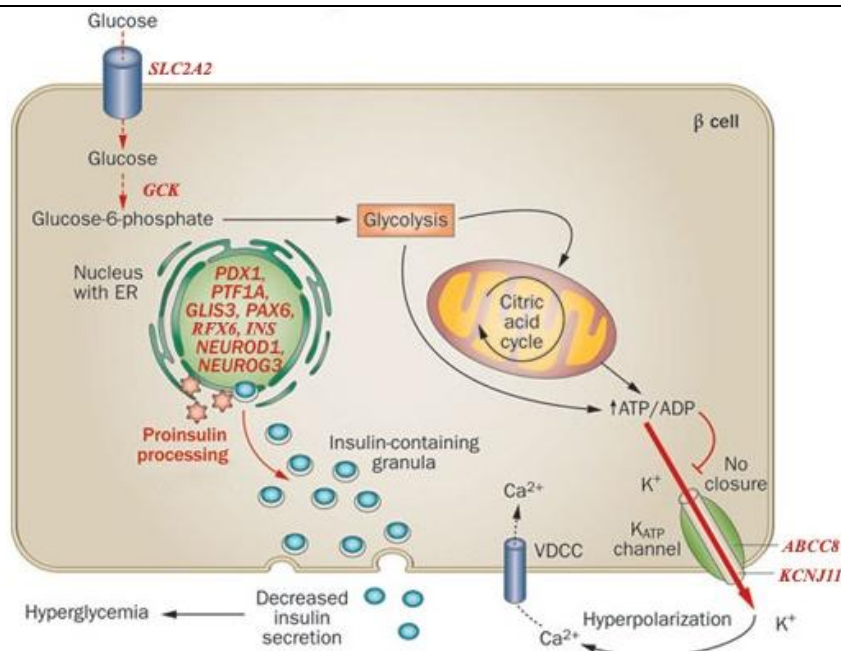


Fig. 15 : Localisation des défauts génétiques dans la cellule  $\beta$  et leurs effets sur la sécrétion d'insuline<sup>209</sup>



## B. L'obésité

### 1) Contexte

Le mot obésité est emprunté au latin *obesitas* qui signifie excès d'embonpoint. Si on décompose le mot *ob* signifie « sur » et *Ēsus* est le participe passé d'*edere* qui signifie manger. L'obésité est un concept ancien qui date de la Grèce antique. Le médecin Hippocrate a été le premier à associer l'obésité comme étant une maladie en faisant l'observation que la mort subite est plus fréquente chez les obèses que chez les minces.<sup>229</sup> Sushruta, un docteur indien du VI<sup>e</sup> siècle av. J.-C., a fait le lien entre obésité, diabète et problèmes cardiaques et il pensait l'exercice physique aiderait à soigner ces 3 maladies.<sup>230</sup> Les conséquences sanitaires de l'obésité ont commencé à être mentionnées dans la littérature médicale du XVIII<sup>e</sup> siècle. William Cullen décrit le syndrome de Pickwick dont les difficultés respiratoires ne seront pleinement documentées qu'en 1956 par C.S.Burwell.<sup>231</sup> Au XIX<sup>e</sup> siècle, William Osler attribue l'obésité à une surconsommation, mais sans mentionner les aspects négatifs de la maladie.<sup>232</sup> À cette époque, 10 à 20kg de plus que le poids idéal était considéré comme une réserve de vitalité en cas de maladie. Il faut attendre 1920 pour qu'une compagnie d'assurance, Metropolitan Life, fasse le lien entre l'obésité et les maladies associées.<sup>233</sup> Le corps médical reconnaît peu à peu l'obésité comme un problème de santé avec l'âge d'or de la psychologie comme la description de la boulimie et de l'anorexie. Dès les années 60, les études sur l'obésité s'intensifient et la graisse est définie comme un organe à part entière avec ses propres hormones, ses récepteurs, sa génétique et sa biologie cellulaire plutôt qu'un simple stockage passif d'énergie.<sup>234</sup> Certaines institutions peinent encore aujourd'hui à reconnaître l'obésité comme une maladie et pas seulement comme une condition ou un trouble. En effet, il faut attendre juin 2013 pour que l'Association médicale américaine (AMA) reconnaisse l'obésité comme une maladie, les enjeux financiers étant extrêmement importants.<sup>235</sup>

### 2) Définition

Dans la seconde moitié du 19<sup>e</sup> siècle, les balances ont permis d'avoir une mesure précise de la masse du corps.<sup>236</sup> La disponibilité grandissante de cette donnée a été la base d'études statistiques qui ont montré un effet différentiel du poids sur la mortalité. C'est Louis I. Dublin, statisticien et vice-président d'une compagnie d'assurance, qui en est à l'origine.<sup>237</sup> Sachant que le poids varie avec la taille, l'indice de Quetelet fit son apparition.<sup>238</sup> Cet indice, connu actuellement sous le nom d'IMC (ou BMI en anglais), est maintenant utilisé pour la classification de l'obésité. Cet indice tire son nom d'Adolphe Quetelet, un astronome belge reconverti en statisticien, qui dans sa tentative de définir l'homme moyen utilisa les données de taille et de poids des armées françaises et écossaises pour montrer que la plupart des cas tombaient dans une fenêtre de valeurs définies par le poids divisé par la taille au carré.<sup>239</sup> L'OMS recommande l'utilisation des normes résumées dans la **Fig. 16** pour les adultes.

L'IMC évalue indirectement l'adiposité. L'absorption biphotonique à rayons X ou scanner DEXA permet une mesure plus précise de l'adiposité, mais des restrictions éthiques et économiques ne permettent pas un usage généralisé de cette méthode. L'IMC reste un indice biaisé par exemple chez les sportifs et certaines ethnies asiatiques. De même, il ne peut pas être utilisé chez les enfants sans quoi près d'un quart d'entre eux ne seraient pas diagnostiqués correctement<sup>240</sup>, et on utilise

pour ces derniers un z-score qui correspond à une mesure corrigée de l'IMC qui prend en compte l'âge et le sexe. Ainsi, un z-score supérieur à 1 correspond à un surpoids et supérieur à 2 à une obésité infantile. Cependant, ces indices et d'autres restent controversés aussi bien chez les adultes<sup>241</sup> que chez les enfants<sup>242</sup>. L'OMS définit l'obésité comme une accumulation anormale ou excessive de graisse qui peut nuire à la santé. Cependant, l'IMC n'est pas forcément le reflet de l'état de santé du patient et de nouveaux indices tenant compte de l'âge, le sexe, la génétique, les caractéristiques cardiométaboliques, les maladies préexistantes et d'autres facteurs semblent essentiels.<sup>243</sup>

Le tissu adipeux est composé de deux types de tissus : le tissu blanc (WAT) qui accumule des lipides et le tissu brun (BAT) qui peut brûler des lipides pour produire de la chaleur. La quantification de ces deux tissus est cruciale pour le développement de nouvelles stratégies thérapeutiques, mais aussi pour la compréhension de la maladie. Le dosage de ces deux tissus n'est pas possible avec un scanner DEXA, mais des recherches récentes ont mis en évidence l'utilisation d'autres techniques d'imagerie telles que le CT-scan<sup>244</sup> et un type d'IRM<sup>245,246</sup>.

Classification	IMC (kg/m <sup>2</sup> )
Insuffisance pondérale	<18,50
Maigreur sévère	<16,00
Maigreur modérée	16,00 – 16,99
Maigreur légère	17,00 – 18,49
Normopondéral	18,50 – 24,99
Surpoids	≥25,00
Préobèse	25,00 – 29,99
Obèse	≥30,00
Obésité classe I	30,00 – 34,99
Obésité classe II	35,00 – 39,99
Obésité classe III	≥40,00

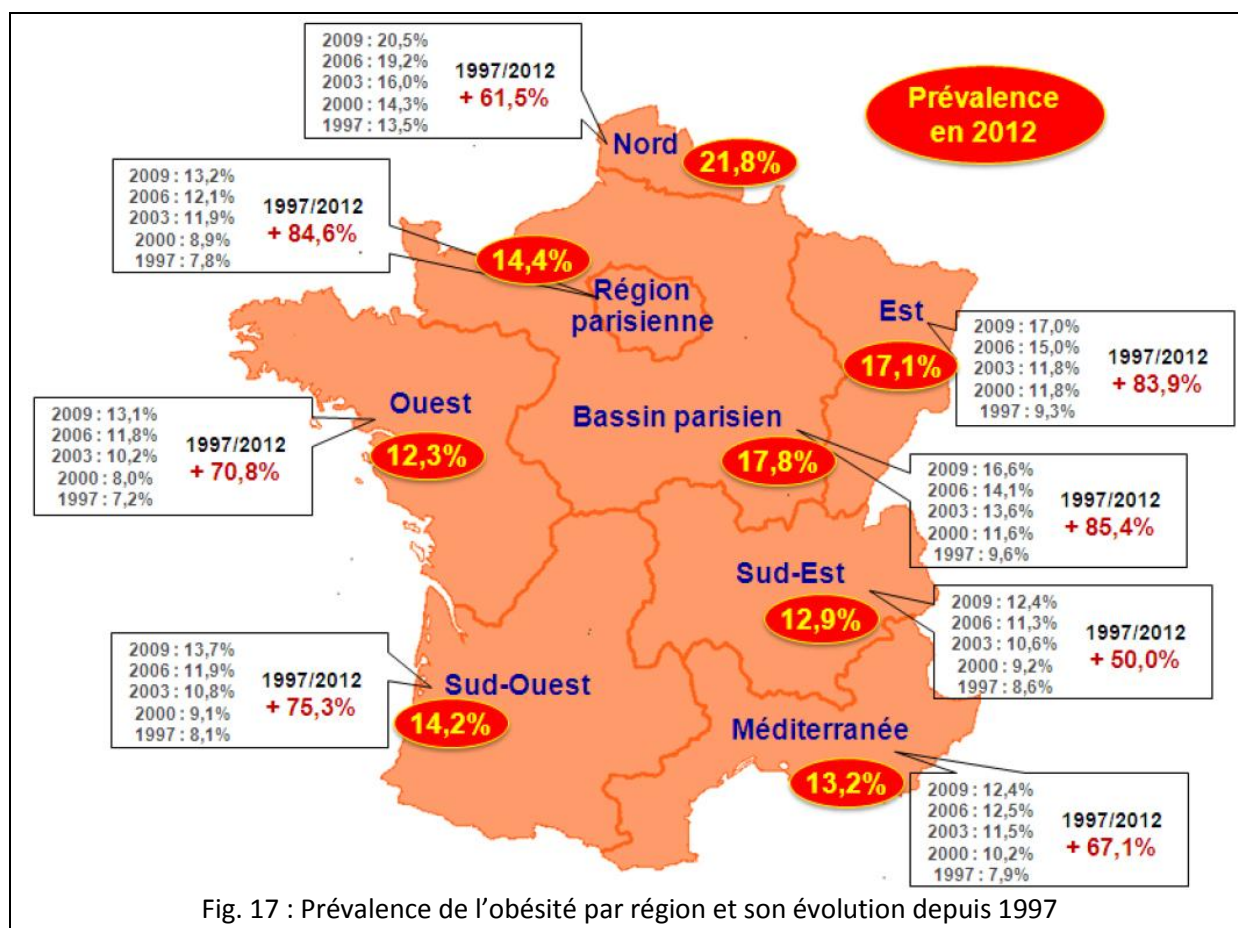
Fig. 16 : Classification internationale de l'obésité adulte

### 3) Épidémiologie

L'augmentation de la prévalence du surpoids et de l'obésité est un problème de santé mondial. En effet, 3,4 millions meurent chaque année à cause de leurs surpoids ou obésité.<sup>247</sup> Une analyse des études épidémiologiques de 183 pays a permis d'estimer à 2,1 milliards le nombre d'adultes en surpoids et obèses dans le monde, soit une augmentation de 27,5% de la prévalence entre 1980 et 2013.<sup>248</sup> Les résultats de cette étude montrent un pic d'obésité entre 1992 et 2002 et principalement chez des personnes âgées entre 20 et 40 ans. L'IMC moyen mondial a augmenté de 0,4kg/m<sup>2</sup> par décennie pour les hommes et de 0,5kg/m<sup>2</sup> par décennie pour les femmes.<sup>249</sup> La prévalence de l'obésité a plus que doublé entre 1980 et 2013 et une étude des Centres pour le contrôle et la prévention des maladies (CDC) a estimé que la proportion d'obèses pourrait atteindre 42% d'ici 2030, soit une augmentation de 33% de la prévalence.<sup>250</sup> La prévalence du surpoids et de l'obésité est plus importante chez les hommes que chez les femmes dans les pays développés et cette tendance s'inverse dans les pays en développement. Parmi les pays à haut revenu, les USA suivis par l'Australie et le Royaume-Uni ont eu la plus forte augmentation de prévalence d'obésité.

Les pays en développement ne sont pas épargnés puisque les taux d'Obésité dépassent 50% dans les Tonga pour les hommes et les femmes, et chez les femmes uniquement au Koweït, en Libye, au Qatar, dans les Kiribati, en Micronésie, et dans les Samoa. D'un point de vue plus global, les États-Unis remportent la palme avec 32% d'hommes obèses et 34% de femmes ; en Europe de l'Ouest, 21% d'adultes sont obèses et seulement 5% des femmes et 8% des hommes en Asie du Sud. On estime que les individus obèses coûtent 30% de plus au niveau des soins de santé qu'un individu de poids normal.<sup>251</sup> Ces frais de santé devraient doubler chaque décennie pour être responsables de plus de 16% des dépenses totales de santé aux USA en 2030, soit environ 860 milliards de dollars.<sup>252</sup>

L'étude Obépi de 2012 rapporte que 32,3% des Français adultes de 18 ans et plus sont en surpoids et 15% présentent une obésité soit environ 7 millions d'obèses.<sup>253</sup> En 15 ans, l'IMC moyen a augmenté de 1,1 kg/m<sup>2</sup>. Parmi les régions UDA8, le Nord est la région à la plus forte prévalence d'obésité en 2012, suivi du bassin Parisien et de l'Est. (Fig. 17) Les 5 autres régions ont des prévalences de l'obésité relativement semblables et toutes inférieures à la moyenne nationale. Les régions subissant les plus fortes évolutions de l'obésité entre 1997 et 2012 sont la région Parisienne, le bassin Parisien et l'Est. On observe donc des disparités interrégionales avec un gradient nord-sud et est-ouest.



Une étude s'appuyant sur les données de l'enquête menée en 2002 par l'IRDES a estimé le coût associé à l'obésité en France.<sup>254</sup> La consommation totale de soins et de biens médicaux d'un individu obèse a été estimée à 2 500 €, 2 fois plus élevée que celle d'un individu de poids normal. Le coût total annuel de l'obésité est estimé entre 2,1 et 3,3 milliards d'euros. Cela représenterait entre 1,5 et 2,4 % de la dépense courante de santé 2002 en France.

#### 4) Complications

L'obésité est un problème de santé publique à cause de sa contribution au développement d'un certain nombre de maladies chroniques qui conduisent à une augmentation de la morbidité et de la mortalité.<sup>255</sup> Les maladies les plus communes pour lesquelles l'obésité est un facteur de risque important sont le diabète de type 2, l'hypertension, le syndrome métabolique, les maladies cardiovasculaires et certains cancers.

Actuellement, on observe une augmentation de la prévalence du diabète de type 2 qui est analogue à l'augmentation de la prévalence de l'obésité.<sup>256</sup> Cela se traduit par une forte corrélation entre le poids moyen et la présence de DT2<sup>257</sup> et une relation inverse entre IMC et l'âge de début de diabète<sup>258</sup>. Concrètement, pour un  $IMC \geq 30$  le risque de développer un DT2 est multiplié par 2, pour un  $IMC \geq 35$  par 5 et pour un  $IMC \geq 40$  par 10.<sup>87</sup> Une étude internationale portant sur 168 000 personnes a démontré une forte association entre le tour de taille et le risque de DT2.<sup>259</sup> Cette étude souligne l'importance de l'adiposité viscérale qui a été démontrée comme précédant le développement du DT2 chez des Américains japonais.<sup>260</sup>

L'obésité est également liée à l'hypertension artérielle : 78% des cas d'hypertension chez les hommes et 65% chez les femmes sont imputés à l'obésité.<sup>259</sup> On observe une relation linéaire entre la tension artérielle et l'IMC à la fois chez les sujets normotendus et hypertendus.<sup>261</sup> Pour chaque augmentation de 10kg par rapport au poids normal, la pression systolique augmente de 3 mmHg et la pression diastolique de 2 mmHg. L'étude de Framingham a montré que sur une période de 4 ans un gain de poids de 5% augmentait le risque d'hypertension de 30%.<sup>262</sup> La répartition de l'adiposité a un impact sur la pression sanguine avec une plus forte association pour l'adiposité viscérale que l'adiposité périphérique.<sup>263</sup> De la même façon que pour le DT2, l'adiposité viscérale permet de prédire l'hypertension chez des Américains japonais.<sup>264</sup>

L'état dyslipidémique est fréquemment observé chez les patients avec une obésité viscérale. Il comprend des niveaux élevés de triglycérides, un faible taux de lipoprotéines de haute densité (HDL), des taux relativement normaux de lipoprotéines de basse densité (LDL) et de cholestérol total, mais un plus grand nombre de particules de LDL qui sont plus petites et plus denses que la normale. Dans un cadre clinique typique, l'hypertriglycéridémie et le faible taux de HDL seront les deux principales anomalies détectables associées à l'obésité. Chaque composante de la dyslipidémie pourrait contribuer au développement de l'athérosclérose. L'hypertriglycéridémie a été proposé comme un facteur de risque indépendant d'athérosclérose, mais cette question reste débattue.<sup>265,266</sup> Les vestiges de lipoprotéines riches en triglycérides, qui sont élevés dans les états hypertriglycéridémiques, sont très athérogènes, probablement autant que les petites particules denses de LDL.<sup>267</sup> La combinaison de taux élevés de triglycérides, de faibles taux de HDL et de petites particules denses de LDL, a été nommée « triade lipidique athérogène », et a été reconnue comme un facteur important de risque cardiovasculaire.<sup>268</sup>

Le syndrome métabolique (MetS) représente un ensemble d'anomalies physiques et métaboliques qui sont des facteurs de risque pour le développement de maladies cardiovasculaires. Un consortium de 6 institutions différentes est finalement parvenu à une définition commune. (**Fig. 18**)<sup>269</sup> La présence de 3 des 5 facteurs de risque constitue un diagnostic de syndrome métabolique.

Mesure	Seuil
Tour de taille	Spécifique de chaque ethnie
Triglycérides	≥150 mg/dL (1.7 mmol/L)
HDL-Cholestérol	<40 mg/dL (1.0 mmol/L) chez les hommes <50 mg/dL (1.3 mmol/L) chez les femmes
Pression sanguine	Systolique ≥130 et/ou diastolique ≥85 mm Hg
Glycémie à jeun	≥100 mg/dL

Fig. 18 : Critères pour le diagnostic clinique du syndrome métabolique

Il a été montré que plus le nombre de facteurs de risque augmente plus le risque de mortalité est élevé.<sup>270</sup> En clinique, le diagnostic du MetS permet d'identifier les individus ayant un risque élevé de développer un DT2 ou une maladie cardiovasculaire. Cependant, cet indice reste controversé dans la mesure où des individus en surpoids ou obèses sans MetS restent à risque pour les maladies cardiovasculaires (CVD).<sup>271</sup> Il semble plutôt que l'adiposité viscérale soit le vrai facteur de risque des CVD et du DT2.<sup>272,273</sup> D'autres problèmes vasculaires sont liés à l'obésité et on peut citer l'accident vasculaire cérébral<sup>274</sup> et l'insuffisance cardiaque<sup>275</sup>.

L'obésité est également un facteur de risque pour plusieurs cancers.<sup>276</sup> (Fig. 19) Les données d'une large analyse prospective ont montré que chez les hommes et les femmes respectivement 14% et 20% des morts de cancer étaient attribuables au surpoids et à l'obésité.<sup>277</sup> Chez l'homme, certains types de cancers sont davantage associés à l'obésité, le cancer colorectal, du pancréas et de la thyroïde.<sup>278</sup> Chez la femme, ce sont plutôt les cancers de l'endomètre, de la vésicule biliaire et des seins qui sont associés à l'obésité.<sup>278</sup>

Type de cancer	Risque relatif avec un IMC de 25-30	Risque relatif avec un IMC ≥ 30	FER (en %) pour la population US	FER (en %) pour la population EU
Colorectal (H)	1,5	2	35,4	27,5
Colorectal (F)	1,2	1,5	20,8	14,2
Seins	1,3	1,5	22,6	16,7
Endomètre	2	3,5	56,8	45,2
Reins	1,5	2,5	42,5	31,1
Œsophages	2	3	52,4	42,7
Pancréas	1,3	1,7	26,9	19,3
Foie	NC	1,5-4	NC	NC
Vésicule biliaire	1,5	2	35,5	27,1
Cardia	1,5	2	35,5	27,1

Fig. 19 : Cancers associés à l'obésité

(NC = non connu, FER = fraction étiologique du risque dans la population)

D'autres complications sont associées à l'obésité : les calculs biliaires<sup>279</sup> et autres maladies de la vésicule biliaire<sup>280</sup>, l'arthrose des genoux et de la hanche<sup>281</sup>, le syndrome d'apnées du sommeil<sup>282</sup>, la dépression<sup>283</sup> et l'hépatopathie stéatosique non alcoolique (NAFLD)<sup>284</sup>. La NAFLD est une maladie où la structure du foie est perturbée par l'accumulation de graisse. La NAFLD peut évoluer de la stéatose à l'hépatite stéatosique non alcoolique (ou NASH) ou même la cirrhose.

## 5) Physiopathologie

La gestion de la quantité d'énergie que nous ingérons implique plusieurs mécanismes et des réseaux qui relient le cerveau à l'intestin, ce processus étant la clé de la régulation du poids corporel au cours du temps, ainsi que la modification des comportements alimentaires à long terme. L'obésité est caractérisée par une augmentation du tissu adipeux sous-cutané. Ses conséquences métaboliques, telles que la résistance à l'insuline, sont principalement dues à une augmentation des dépôts de graisse sur des sites tels que l'épiploon, le foie et les muscles squelettiques.<sup>285</sup>

D'un point de vue pratique, il n'y a que deux façons de contrôler le poids d'un individu : l'apport alimentaire et la quantité d'activité physique. Du point de vue de la thermodynamique, les changements de poids sont la résultante de la balance énergétique : le gain de poids se produit lorsque l'apport énergétique est supérieur à la dépense énergétique, et la perte de poids se produit lorsque la dépense d'énergie est supérieure à l'apport énergétique. L'énergie est dépensée par trois voies principales : 1) le taux métabolique au repos (RMR), qui est l'énergie dépensée pour exercer des fonctions corporelles au repos, incluant la réparation et la maintenance des cellules et des tissus, le maintien de la fréquence cardiaque, une température constante du corps, et d'autres processus physiologiques nécessaires au maintien d'un milieu environnemental et hormonal interne constant, 2) l'effet thermique des aliments, ce qui correspond à l'excès de chaleur générée quand la nourriture est consommée, digérée et absorbée, et 3) l'activité physique, qui est variable et la seule composante de la dépense énergétique qui est soumise à un contrôle volontaire.<sup>286</sup> La composante la plus importante de la dépense d'énergie est le RMR, qui représente généralement 60-70% de la dépense énergétique quotidienne totale (TEE).<sup>287</sup> L'effet thermique des aliments représente environ 10 % de la TEE. La dépense énergétique liée à l'activité physique dépend de la quantité et de la nature de l'activité physique. Chez une personne très active, l'énergie dépensée par l'activité physique peut atteindre 40 % de la TEE, mais elle peut être à seulement 10 % de la TEE chez un individu sédentaire.

On sait maintenant que l'un des composants clés qui contrôlent la balance énergétique est l'hypothalamus. Cette découverte date de la fin du 19e siècle avec la description du syndrome de Fröhlich caractérisé par des tumeurs pituitaires associées à un hypogonadisme et une accumulation sous-cutanée excessive de graisse.<sup>288,289</sup> Toutefois, on ne savait pas si ce syndrome était dû à la glande pituitaire ou à cause de dommages sur l'hypothalamus. Plus tard, Hetherington et Ranson démontrèrent chez le rat que la destruction de l'hypothalamus médial sans dommage de la glande pituitaire causait une obésité morbide et des dérèglements neuroendocrines similaires à ceux du syndrome de Fröhlich.<sup>290</sup> Si le rôle de l'hypothalamus était alors bien établi, on ne connaissait pas les mécanismes moléculaires sous-jacents qui impliquaient un dialogue avec le tissu adipeux. Il faut attendre 1950 pour découvrir des mutations récessives *ob* et *db* qui rendent les souris hyperphagiques et obèses morbides.<sup>291</sup> Les gènes *ob* et *db* ont été clonés vers le milieu des années 90.<sup>292-294</sup> L'hormone issue du gène *ob* a été appelée leptine, qui vient du grec *leptos* signifiant mince, car son injection conduit la souris à un état cachectique (amaigrissement extrême).<sup>295-298</sup> De la même façon, le clonage du gène *agouti* issu des souris jaune *agouti*<sup>299-301</sup> a conduit à la caractérisation du système de la mélanocortine centrale<sup>302,303</sup>.

L'obésité survient à la suite d'un déséquilibre entre l'apport et la dépense énergétique régulés autour par un axe composé du cerveau, des intestins et du tissu adipeux. (**Fig. 20**) De



nombreuses informations sont déjà disponibles concernant les gènes, les peptides, les neurotransmetteurs et les récepteurs dans l'hypothalamus et les régions voisines qui régulent l'appétit et le poids corporel. Le neuropeptide principal qui augmente l'appétit est le neuropeptide Y. Les autres neuropeptides augmentant l'appétit sont l'orexine A et B, l'agouti related peptide (AgRP) et l'hormone qui concentre la mélanine (MCH). Les neuropeptides qui diminuent l'appétit sont les dérivées de la pro-opiomélanocortine (POMC) qui agissent sur le récepteur MC4, le peptide CART, la corticolibérine (CRH), le prolactin releasing peptide (PrRP), la mélanocortine alpha ( $\alpha$ -MSH), la sérotonine (5HT) et le récepteur à la leptine (LEPR).

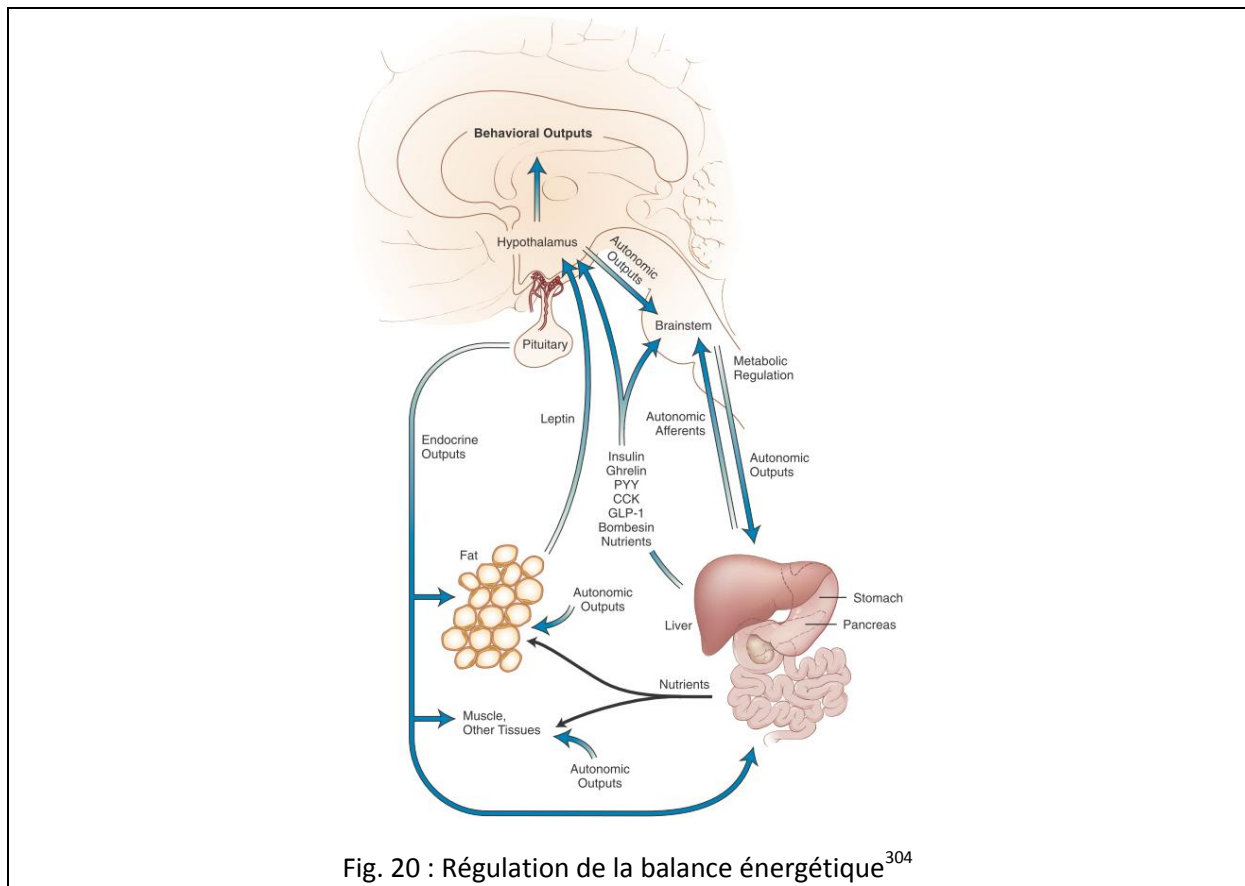


Fig. 20 : Régulation de la balance énergétique<sup>304</sup>

Il y a 4 hypothèses principales concernant les mécanismes afférents impliqués dans la régulation de l'appétit. Selon l'hypothèse lipostatique, le tissu adipeux produit un signal hormonal qui est proportionnel à la quantité de graisse et qui agit sur l'hypothalamus pour diminuer la prise alimentaire et augmenter la dépense énergétique.<sup>305</sup> Cette hypothèse est appuyée par le fait que le tissu adipeux blanc libère la leptine et la résistine qui diminuent l'appétit. La condition inflammatoire de l'organisme due à l'obésité serait aussi liée à cette hypothèse. Par exemple, la libération d'adiponectine et d'adipocytokines tels que le facteur de nécrose tumorale et l'interleukine-6 augmentent aussi l'appétit. Le tissu adipeux brun libère également PPAR- $\gamma$  et la thermogénine (UCP-1) qui sont responsables d'un taux métabolique élevé et donc de la réduction du poids.

L'hypothèse du peptide gastro-intestinal met en jeu certaines hormones du système digestif comme la bombésine (GRP) sécrétée par l'estomac. On peut également citer le glucagon et la somatostatine sécrétés par le pancréas qui diminuent l'appétit et contrôlent le poids, les peptides CCK et PYY sécrétés par l'intestin et le colon qui sont aussi responsables de la régulation de l'appétit

et du poids, et la ghréline sécrétée par l'estomac qui exerce un effet orexigène par la voie NPY/AGRP dans le noyau arqué.

L'hypothèse glucostatique stipule qu'une diminution du taux de glucose sanguin augmente l'appétit<sup>306</sup>, des jeûnes fréquents conduisant à la réduction du métabolisme de base et une augmentation de l'adiposité. L'hypothèse sous cette forme a été abandonnée. La notion de détection du glucose dans le cerveau a été mise en lumière avec la découverte des neurones senseurs de glucose dans l'hypothalamus, y compris les neurones exprimant NPY et POMC dans le noyau arqué, qui sont également sensibles à la leptine, à l'insuline et à d'autres neuropeptides.<sup>307,308</sup> Il a été émis l'hypothèse que le glucose, l'insuline et la leptine agissent dans ces neurones par l'intermédiaire de la glucokinase et de la PI3-kinase afin de stimuler le métabolisme du glucose intracellulaire, qui fournit de l'ATP et à son tour réduit l'activité de l'AMPK et la prise alimentaire.<sup>308,309</sup>

L'hypothèse thermostatique postule que la chute de la température corporelle en dessous d'un certain point stimule l'appétit et au-dessus de ce même point l'inhibe.<sup>310</sup> Cette hypothèse en l'état est abandonnée.

## 6) Traitements

### a) L'approche comportementale

Que ce soit pour la perte de poids ou pour prévenir le gain de poids ou son regain, la thérapie comportementale est une approche primaire d'intervention de l'obésité. En effet, le National Heart, Lung, and Blood Institute (NHLBI) recommande comme premier traitement de l'obésité un programme composé de régime alimentaire, d'exercice et de thérapie comportementale.<sup>311</sup>

Le traitement comportemental est un ensemble de plusieurs composantes destiné à faciliter les changements d'alimentation et d'activité physique. Ces procédures ont été décrites dans les manuels de traitement détaillés, tels que le programme pour la gestion du poids LEARN<sup>312</sup> et le programme de prévention du diabète DPP<sup>313</sup>. Les composantes de ce traitement sont l'autosurveillance (apport alimentaire, activité physique et suivi du poids), le contrôle du stimulus via le conditionnement classique et opérant (par exemple en dissociant l'action de manger et regarder la télé), la résolution de problèmes (comprendre les chaînes d'événements qui conduisent à trop manger ou à peu d'activité physique), la restructuration cognitive (éviter les baisses de moral dues à un seul excès alimentaire), et la prévention des rechutes (anticiper les périodes à risque telles que les vacances ou au contraire les périodes de stress au travail).

De façon générale, les patients ayant reçu la thérapie comportementale ont perdu en moyenne environ 10% de leur poids initial (l'équivalent de 10 kg) avec en moyenne 30 semaines de traitement. (**Fig. 21**) Dans les années 90, la durée de traitement augmente sensiblement, ce qui a été associé à une augmentation correspondante de la perte de poids. Les participants ont généralement perdu 0,4-0,5 kg par semaine<sup>314</sup> et ont atteint leur perte de poids maximal à six mois<sup>315</sup>. Le taux d'arrêts prématurés du traitement était d'environ 20 %. De nombreuses études ont effectué des évaluations de suivi des adultes qui ont participé à des programmes de traitement de comportement. En moyenne, 30-35% de la perte de poids ont été retrouvés dans l'année suivant le traitement.<sup>316</sup> Entre 4 et 5 ans après le traitement, les patients ont maintenu une perte de poids



moyenne d'environ 1-3 kg, et de nombreux patients ont repris la totalité du poids qu'ils ont perdu.<sup>317,318</sup>

<b>Caractéristiques</b>	<b>1974</b>	<b>1985-87</b>	<b>1991-95</b>	<b>1996-2002</b>
Nb d'études	15	13	5	9
Taille moyenne de l'échantillon	53,1	71,6	30,2	28
Poids initial (en kg)	73,4	87,2	94,9	92,2
Durée du traitement (en semaines)	8,4	15,6	22,2	31,4
Perte de poids (en kg)	3,8	8,4	8,5	10,7
Perte par semaine (en kg)	0,5	0,5	0,4	0,4
Abandons (en %)	11,4	13,8	18,5	21,2
Durée du suivi (en semaines)	15,1	48,3	47,7	41,8
Perte de poids à la fin du suivi	4	5,3	5,9	7,2

Fig. 21 : Résumé de l'efficacité du traitement comportemental sur l'obésité<sup>314</sup>

Ces résultats sont décevants pour la plupart des patients. Cependant, de nouvelles interventions ont facilité une plus grande réussite à long terme, tel que celles intégrant un modèle de soins continus après la perte de poids initiale. Les données publiées de l'étude Look AHEAD constatent que, après quatre ans de traitement continu, les participants ont maintenu une perte de poids de 4,7% du poids initial.<sup>319</sup> Le programme DPP a également démontré que des pertes de poids relativement modestes produisent des bénéfices significatifs pour la santé.<sup>77</sup> Dans cette étude, même dans un contexte de reprise de poids, les participants qui ont reçu le traitement comportemental ont réduit leur risque de développer le diabète de type 2 de 58% par rapport aux participants sous placebo.

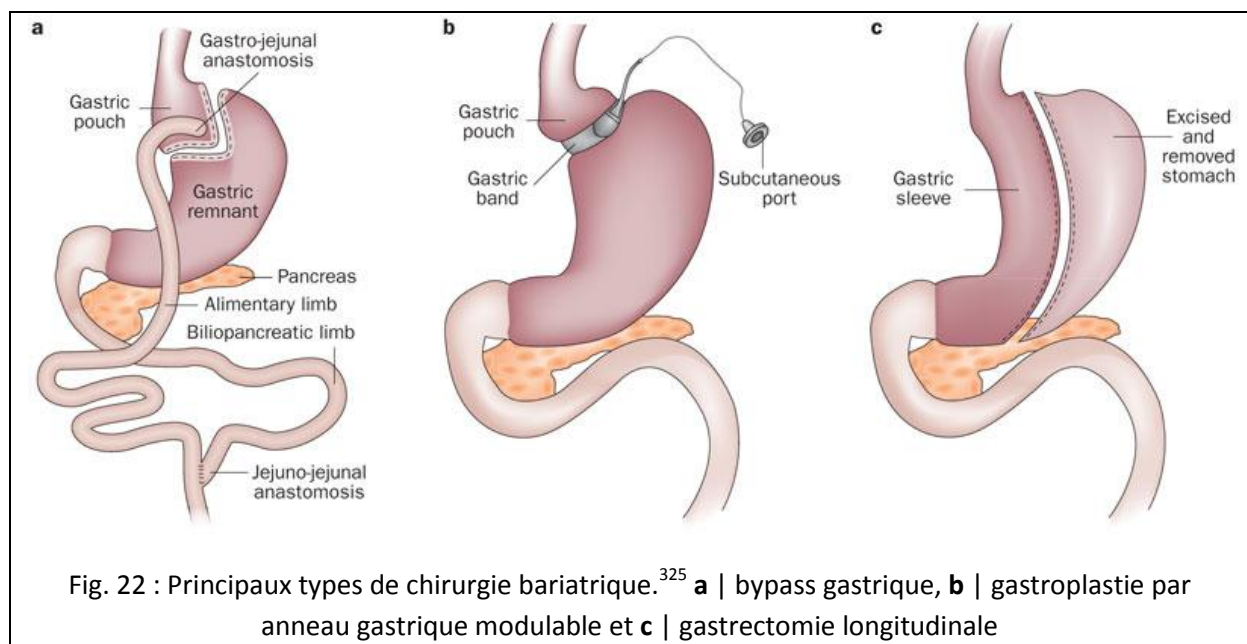
Dans ce type d'approche, la question du type de régime se pose. Une méta-analyse de cinq études avec 6 et 12 mois de suivi n'a révélé aucune différence dans la perte de poids entre les régimes à faible teneur en glucides et ceux faibles en graisses.<sup>320</sup> Le taux de cholestérol et de LDL s'est amélioré sur les régimes faibles en graisses, alors que les valeurs de HDL et de triglycérides étaient meilleures pour les régimes faibles en glucides. Une étude de deux ans comparant le maintien à long terme d'un régime à faible teneur en graisses, méditerranéen, et ceux à faible teneur en glucides a trouvé un meilleur maintien sur le régime faible en glucides pour les six premiers mois, mais pas à deux ans.<sup>321</sup> Le maintien du régime faible en glucides a été considérablement réduit et le taux d'abandons était le plus élevé des trois régimes, tandis que ceux qui suivaient un régime méditerranéen ont démontré un fort maintien et des taux d'abandon faibles. Les régimes qui sont plus restrictifs dans la diversité pourraient réduire le nombre de calories consommées tout simplement parce que les choix disponibles sont limités, mais sont moins susceptibles d'être suivis sur le long terme.

#### b) La chirurgie bariatrique

Une modification du mode de vie sous la forme d'un régime hypocalorique peut être très efficace pour la perte de poids, mais 80-90% de la population ne parvient pas à maintenir la perte de poids à long terme en raison de processus de compensation.<sup>322</sup> La chirurgie bariatrique reste le traitement le plus efficace pour la perte de poids et son maintien à long terme chez les obèses morbides.

Les premières interventions de chirurgie bariatrique ont été développées dans les années 1950. À l'époque, les chirurgiens ont tenté de favoriser la perte de poids grâce à l'élaboration de procédures que restreint l'apport alimentaire ou conduisent à la malabsorption de calories. Le bypass gastrique Roux-en-Y (RYGB), la gastroplastie par anneau gastrique modulable (AGB) et la dérivation bilio-pancréatique (avec ou sans switch duodénal) sont les trois techniques qui ont résisté à l'épreuve du temps, et ce sont certainement les deux premières qui continuent d'être le plus utilisées à travers le monde. L'augmentation de l'utilisation de la chirurgie bariatrique est sans précédent : entre 2000 et 2010, le nombre d'opérations a été multiplié par 10 en Angleterre.<sup>323</sup>

Aujourd'hui, il existe deux types principaux d'interventions, restrictives et combinées. (**Fig. 22**) Les premières visent à réduire la capacité gastrique, c'est-à-dire le volume utile de l'estomac (par ex. la gastrectomie longitudinale ou VSG) ou à réduire la vitesse de vidange de l'estomac afin d'obtenir un sentiment de satiété plus rapidement (par ex. anneau gastrique ou AGB). Les secondes associent à cette restriction gastrique la création d'un système de dérivation dans le tube digestif afin de diminuer l'absorption des éléments nutritifs par l'intestin, notamment par la technique du bypass gastrique (RYGB). Le bypass gastrique agit en réduisant la faim, en augmentant la satiété, en changeant les préférences alimentaires et en augmentant la dépense énergétique. L'AGB fonctionne probablement par la réduction de la faim qui pourrait être médiée par la signalisation vagale. Si la VSG réduit le volume de l'estomac, son action principale sur la perte de poids serait plutôt due à une augmentation des acides biliaires et des changements de la flore intestinale.<sup>324</sup>



La RYGB, l'AGB et la VSG sont les procédures les plus couramment effectuées et entraînent une perte de poids de 20-35% et son maintien.<sup>326,327</sup> Les améliorations cliniques dans les comorbidités associées à l'obésité, notamment les maladies métaboliques, cardiovasculaires, respiratoires, musculo-squelettiques et rénales, le cancer, mais aussi la réduction de la mortalité globale et cardiovasculaire associée, peuvent être considérables.<sup>328-333</sup> De plus, deux études parues dans le New England Journal of Medicine (NEJM) ont montré que la chirurgie bariatrique permettait un meilleur contrôle de la glycémie avec plus de 75% de rémission à 1 an et plus de 90% à 3 ans chez les patients DT2 par rapport aux traitements classiques<sup>85,86,334</sup>, et une autre étude a montré que le

risque de développer un DT2 diminuait de 80%<sup>335</sup>. Cependant, les individus opérés présentent une augmentation des symptômes dépressifs<sup>336</sup>, des risques de maux de tête sévères<sup>337</sup>, et une mortalité par accident ou par suicide accrue de 58%<sup>338</sup>. Une partie non négligeable des patients obèses a été victime de violence durant leur enfance<sup>339,340</sup>, et il est possible que ce traumatisme réapparaisse après la chirurgie.

### c) Pharmacothérapie

Compte tenu de l'échec de l'approche comportementale et les limites de la chirurgie bariatrique, l'utilisation de médicaments anti-obésité comme traitement d'appoint dans le traitement de l'obésité reste essentielle. L'objectif est de développer un schéma thérapeutique sûr et efficace, en combinaison avec une meilleure alimentation et de l'exercice, pour atteindre une réduction significative et durable du poids corporel et de profiter des avantages qui en découlent. Un médicament pour la perte de poids doit être compatible avec le profil d'un patient obèse pour être vraiment significatif en termes d'efficacité, de sécurité et de durabilité. En effet, la population obèse est très hétérogène, avec des variations de degré et de durée de la surcharge pondérale, d'âge et des comorbidités associées.

L'histoire des médicaments favorisant la perte de poids révèle que les approches pharmaceutiques ont longtemps été utilisées, mais sont également en proie à de sérieux problèmes de sécurité. Dans les années 1890, des extraits de thyroïde de mouton ont été utilisés pour augmenter le métabolisme et induire une perte de poids chez les patients euthyroïdiens, mais ce traitement était aussi responsable d'arythmies et d'arrêts cardiaques.<sup>341</sup> Dans les années 1930, le 2,4-dinitrophénol était un médicament pour la perte de poids très efficace et populaire jusqu'à la découverte d'effets secondaires tels qu'une agranulocytose, des cataractes, une dermatite et une hyperthermie fatale.<sup>342</sup> Durant les années 1940-60, l'usage généralisé de cocktails incluant des amphétamines révéla leurs propriétés addictives ainsi qu'une toxicité myocardique et une mort subite chez de nombreux patients.<sup>343</sup> Des analogues d'amphétamines font ensuite leurs apparitions dont le très populaire Fen-Phen en 1992 qui combine la fenfluramine avec la phentermine. Ce médicament causait une hypertension pulmonaire et une valvulopathie, menant à son retrait et à ses dérivés en 1997, ainsi que des milliers de poursuites et des milliards de dollars de dommages et intérêts.<sup>344</sup> En 2008, l'agence américaine des produits alimentaires et médicamenteux (FDA) et l'agence européenne des médicaments (EMA) ont rejeté le rimonabant, un antagoniste des récepteurs CB1 des cannabinoïdes, en raison de risques psychiatriques, dont la dépression et le suicide.

Actuellement, les options pharmacothérapeutiques sont très limitées. Seulement trois médicaments sont approuvés par la FDA comme traitement d'appoint pour la gestion de poids chronique: l'orlistat (Alli, GlaxoSmithKline, Xenical, Roche) approuvé en 1999, la lorcasérine (Belviq, Arena Pharmaceuticals) approuvée en 2012, et le combiné phentermine / topiramate en formulation à libération prolongée (Qsymia, Vivus) également approuvé en 2012. Aujourd'hui, un seul médicament, l'orlistat, a été approuvé par la FDA et l'Agence européenne des médicaments (EMA) pour la gestion du poids. L'orlistat inhibe les lipases gastro-intestinales, réduisant l'absorption des graisses. Par conséquent, l'effet indésirable le plus fréquent est la stéatorrhée. Malgré son statut

approuvé, l'orlistat a eu un certain nombre d'effets indésirables, incluant l'hépatotoxicité, la néphrotoxicité, la pancréatite, et des calculs rénaux.<sup>345</sup>

Concernant les 2 autres médicaments, la lorcaserine et le combiné phentermine/topiramate, ils ont été rejetés par la FDA en 2010 à cause de problème d'efficacité et de sécurité pour être finalement acceptés en 2012 sous couvert d'études supplémentaires démontrant l'absence d'incidents cardiovasculaires. Pour sa part, l'EMA a rejeté la lorcaserine, car les bénéfices du médicament n'étaient pas supérieurs aux risques, notamment avec un risque potentiel pour les tumeurs,<sup>346</sup> et le combiné phentermine/topiramate en raison de préoccupations quant aux effets potentiels sur le système cardiovasculaire et nerveux central associés à son utilisation à long terme, son potentiel tératogène, et son utilisation sur des patients pour lesquels il n'est pas indiqué.<sup>347</sup>

La nécessité de nouvelles approches pour traiter et prévenir l'obésité et de ses séquelles associées a conduit à plusieurs axes de développement ciblant :

- La signalisation neuropeptidique centrale : agonistes de MCH-R, MC4R, Y4R ou Y5R
- La neurotransmission des monoamines : inhibiteurs de la recapture de la noradrénaline et de la dopamine
- La signalisation peptidique intestinale : agonistes de CCK-A ou GLP-1R
- La signalisation pancréatique : amyline ou analogue du PP, inhibiteur de la lipase pancréatique
- La signalisation du tissu adipeux : agoniste du récepteur de la leptine

Une autre piste sérieuse implique le microbiote intestinal. En effet, les microorganismes de l'intestin sont liés au métabolisme, et il a été suggéré que la flore pourrait avoir un rôle dans l'obésité. Une étude récente a alimenté le débat sur le rôle des bactéries intestinales dans l'obésité en montrant que le microbiote de personnes minces ou obèses peut transmettre les phénotypes de minceur ou d'obésité lorsqu'ils sont transplantés dans des souris.<sup>348</sup> De plus, ces auteurs ont montré que des souris qui ont un microbiote «maigre» ne développent pas de phénotype métabolique associé à l'obésité lorsqu'on leur transplante un microbiote «obésité» issu d'autre souris. Ces résultats intrigants indiquent qu'il pourrait être possible d'inverser l'obésité en modulant la composition du microbiote.

Quand on parle de microbiote, on pense souvent au microbiote intestinal. Pourtant, l'estomac a peut-être aussi un rôle dans l'obésité. En effet, près de 50% de la population est infecté par *Helicobacter pylori* qui entraîne des ulcères à l'estomac chez 20% des sujets. Les patients traités pour cette bactérie ont tendance à grossir.<sup>349</sup> En effet, les taux d'obésité et de surpoids sont inversement corrélés avec la prévalence de l'infection à *H. pylori*.<sup>350</sup> Sachant que le taux d'infection a diminué lors des dernières décennies, il est possible que la bactérie ait un rôle à jouer dans la pandémie d'obésité. Toutefois, cet effet pourrait être indirect en agissant sur la satiété<sup>351,352</sup> ou encore sur le microbiote intestinal<sup>353</sup>.

D'autres pistes impliquent certains fruits. Une étude comparant différents types de baies a montré que le cassis, la myrtille et plus particulièrement les airelles prévenaient partiellement ou complètement la prise de poids chez la souris nourrie à un régime riche en graisses.<sup>354</sup> Les airelles amélioraient même la glycémie et le taux de cholestérol. Le jus de pamplemousse diminuerait

également la prise de poids et la résistance à l'insuline chez des souris soumises à un régime riche en graisses.<sup>355</sup>

## 7) Facteurs de risque

La mondialisation a eu des conséquences telles que les transitions nutritionnelles, mais aussi des changements de modes de vie et du cadre de vie qui sont le moteur de l'épidémie mondiale d'obésité. La libéralisation du commerce et les politiques concernant l'agriculture ont influencé la disponibilité de certains aliments. Par exemple, si les aides gouvernementales favorisent l'agriculture du maïs et du soja, les aliments transformés ainsi que le prix du bétail baisseront, car ils en dépendent. À l'inverse, la production de fruits et légumes qui ne reçoit aucune subvention aura un prix élevé. Ce prix élevé des fruits et légumes est justement corrélé avec un IMC plus élevé chez les enfants.<sup>356</sup> Kenneth Rogoff, un économiste américain, suggère que l'épidémie d'obésité est la conséquence inévitable d'un système alimentaire capitaliste non réglementé, dans lequel les entreprises se concurrencent les unes avec les autres pour inciter les consommateurs à manger plus de leurs produits. Lobstein a calculé que pour réduire l'IMC au Royaume-Uni de 8%, soit le niveau observé en 1980, cela coûterait à l'industrie alimentaire 8,7 milliards de livres sterling par année.<sup>357</sup> L'industrie alimentaire doit donc protéger ses profits et développe des stratégies pour ce faire. Les grosses compagnies alimentaires commercialisent des produits faibles en calories, mais continuent de commercialiser des produits hauts en calories pour multiplier les profits, une hypocrisie qui contribue probablement à l'épidémie d'obésité.<sup>358</sup>

La prévalence de l'obésité corrèle positivement avec les stades initiaux de la croissance économique qui impliquent des changements nutritionnels et de mode de vie. Une analyse mettant en relation les risques nutritionnels avec le développement économique dans 100 pays a montré que l'IMC augmente rapidement avec le revenu national.<sup>359</sup> Cette association est plus faible dans les pays à revenus moyens supérieurs et hauts, à cause d'un meilleur accès aux soins de santé et à l'éducation, mais aussi par des changements de comportement.<sup>359</sup> Dans les pays à revenus moyens et faibles, l'IMC est positivement corrélé avec le statut socioéconomique, à l'opposé des pays à hauts revenus où l'IMC est inversement associé au statut socioéconomique. Une étude a permis de montrer que ce changement arrivait chez les femmes si le PNB par habitant était supérieur à 2500\$.<sup>360</sup>

La proportion de la population mondiale vivant en milieu urbain est passée de 13% au début du 20e siècle à 29% dans les années 1950 pour dépasser maintenant les 50%. On estime que cette proportion devrait augmenter à plus de 70% en 2050, avec une augmentation plus importante dans les pays à revenus moyens et faibles.<sup>361</sup> En Chine par exemple, on estime à plus d'un milliard le nombre de personnes vivant en zone urbaine en 2050. Une étude indienne a montré que la migration vers les zones urbaines était associée à l'obésité.<sup>362</sup> Les conséquences de la vie urbaine sur le développement de l'obésité sont nombreuses et se produisent principalement en raison de changements de style de vie et des choix alimentaires possibles ainsi que des modifications liées au progrès technologique et à la mécanisation. Collectivement, ces changements ont un effet direct sur l'augmentation de l'apport calorique et la diminution de la dépense énergétique qui contribuent au déséquilibre de la balance énergétique.

Les zones densément peuplées avec peu d'espace récréatif extérieur limitent les possibilités pour la marche et d'activité physique concernant les loisirs. De modestes avantages de l'activité physique sur le poids ont été démontrés, avec des effets plus notables lorsque l'activité physique est combinée avec une intervention diététique.<sup>363</sup> Le niveau d'activité physique lié au travail, au transport et aux tâches ménagères a diminué de façon importante et des comportements sédentaires tels que la télévision<sup>364</sup> et l'utilisation d'ordinateurs se sont généralisés diminuant l'activité physique totale.<sup>365</sup> En analysant les données NHANES des 20 dernières années, une étude a montré que le nombre d'Américaines qui déclaraient ne pas faire d'activité physique a augmenté de 19,1% à 51,7% et le nombre d'Américains de 11,4% à 43,5%.<sup>366</sup> L'origine ethnique est un facteur important puisque les noirs américains et les Américains d'origine mexicaine sont ceux qui font le moins d'activité physique. Cette diminution de l'activité physique diminuerait le nombre de calories brûlées, mais aussi la capacité de l'organisme à oxyder les graisses. Une étude a montré une augmentation jusqu'à 20% de l'oxydation des graisses chez des personnes auparavant non sportives et soumises à un programme de remise en forme de 12 semaines.<sup>367</sup>

L'urbanisation contribue aussi au développement de l'obésité par la diminution du temps de sommeil. En zone urbaine plus qu'en zone rurale, la durée de sommeil est plus courte à cause de la pollution sonore, l'éclairage public, l'accès à la télévision et à internet, le travail posté et les activités sociales nocturnes. La durée du sommeil a été associée à l'obésité dans tous les groupes d'âge différents, mais elle est plus forte chez les enfants et les adolescents.<sup>368</sup> Le lien direct n'est pas établi, mais on pense qu'une durée trop courte de sommeil a une influence sur le métabolisme des glucides, la fonction endocrine, la sécrétion de ghréline et de leptine.<sup>369,370</sup> Une autre hypothèse est la durée d'exposition à la lumière matinale. En effet, une étude récente rapporte que plus l'exposition journalière à la lumière a lieu tôt, plus l'IMC est bas.<sup>371</sup> Cette association est indépendante de l'activité physique, de l'apport calorique, de la durée du sommeil, de l'âge ou de la saison. La lumière permet de synchroniser l'horloge interne et joue un rôle dans la régulation du métabolisme, de la faim et de la satiété. Le mode de vie moderne impose souvent à l'intérieur où l'intensité de la lumière n'est pas assez élevée (environ 200 à 300 lux) en comparaison avec l'extérieur où la luminosité est d'au moins 1000 lux même par temps couvert. Cette exposition à la lumière expliquerait jusqu'à 20% de la variance de l'IMC. Enfin, la dernière hypothèse concerne la température des habitats qui a augmenté de 19°C à 22°C en 3 décennies.<sup>372</sup> En effet, une étude récente publiée dans Diabetes démontre que la quantité de graisse brune augmente ou diminue selon la température environnante.<sup>373</sup>

Durant les dernières décennies, le choix des aliments disponibles a drastiquement changé. Par exemple, le nombre de fast foods a explosé tel que le nombre global de McDonald's qui est passé de 951 en 1987 à 7135 en 2002.<sup>374</sup> Une étude menée dans 26 pays riches a montré que les pays ayant une plus grande densité de fast foods par habitant avaient des taux d'obésité beaucoup plus élevés par rapport aux pays ayant une faible densité de fast foods par habitant.<sup>375</sup> De plus, il y a une association positive entre la proximité aux fast foods et l'obésité chez les jeunes enfants.<sup>376</sup> À l'inverse, la proximité aux commerces vendant des aliments sains diminue de 38% le risque d'être obèse ou en surpoids chez les jeunes enfants.<sup>377</sup> En fait, la consommation de fast foods serait un sous-produit d'un problème beaucoup plus important: de mauvaises habitudes alimentaires dues à l'éducation des parents.<sup>378</sup> En effet, les enfants issus de parents divorcés ont 54% de risque d'être en surpoids ou obèses ce qui pourrait être dû à une diminution du temps pour les tâches ménagères telles que la cuisine ou encore une diminution des revenus.<sup>379</sup> Cependant, les fast foods sont souvent



pointés du doigt, mais ils n'expliquent qu'une petite part de l'obésité. Une étude de la répartition spatiale de la propagation de l'obésité suggère que l'obésité serait liée au comportement collectif plus qu'aux choix individuels. En effet, ils ont observé une corrélation entre la géographie de l'épidémie et la commercialisation des aliments et des modes de distribution.<sup>380</sup>

L'autre changement alimentaire important et récent a été l'augmentation du nombre de grands supermarchés qui a supplanté rapidement les marchés des produits alimentaires frais et magasins à la ferme.<sup>381</sup> L'effet des supermarchés sur l'obésité pourrait être important, car ils sont une source d'aliments hautement transformés, de grignotines (snack), bonbons et boissons sucrées ou tout simplement parce que les portions sont trop grandes avec par exemple 53% d'augmentation de la taille des plats tout préparés.<sup>382</sup> D'ailleurs, la capacité à gérer ou à estimer le nombre de calories que nous mangeons devient de moins en moins précise au fur et à mesure que la taille du repas augmente.<sup>383</sup> Chez les adultes, plus la taille de la portion alimentaire est grande plus la quantité de nourriture absorbée est grande.<sup>384</sup> Chez les enfants qui ont pourtant une capacité innée à autoréguler leur apport alimentaire, cette limite peut aussi être outrepassée par la présentation de grandes portions alimentaires qui les encouragent à manger même sans avoir faim.<sup>385</sup> Ce surdimensionnement alimentaire est une stratégie marketing qui offre une valeur ajoutée pour une augmentation minime du prix qui donne au consommateur l'impression d'avoir fait une affaire. D'autres éléments marketing influencent la quantité d'aliments que nous consommons : la saillance du produit par son packaging, la structure par les différentes couleurs du produit, l'empilement des produits, la forme et la taille des contenants, et la publicité télévisée<sup>386</sup>.

En outre, les produits alimentaires sont soigneusement formulés par les fabricants pour maximiser leur attirance aux consommateurs.<sup>387</sup> L'OMS a identifié la palatabilité comme l'une des influences les plus importantes sur le comportement alimentaire, qui agit en augmentant la vitesse à laquelle on mange, ainsi que la sensation de faim pendant et entre les repas.<sup>388</sup> La contribution majeure à la palatabilité des aliments est la graisse, qui offre une agréable sensation en bouche et est un excellent véhicule des arômes, c'est pourquoi elle est utilisée dans les sauces et la crème glacée. Cependant, la graisse contient deux fois plus de calories par gramme que les glucides ou les protéines, et contribue de façon majeure à la densité énergétique des aliments. C'est peut être un facteur favorisant une balance énergétique positive, car si un aliment à haute densité énergétique est consommé rapidement, un grand nombre de calories peut être consommée avant que la satiété ait une chance de se développer.<sup>389,390</sup> La graisse et le sucre sont une combinaison particulièrement savoureuse, que l'on trouve dans les cupcakes, les cookies, etc. Cette combinaison est très attrayante pour les humains, et peut-être plus attrayante encore pour les femmes obèses que de leurs homologues minces.<sup>390</sup> La graisse et le sel est aussi une combinaison très agréable au goût qui interagit avec les récepteurs du goût sur la langue et stimule les signaux neuronaux dans le cerveau qui encouragent la consommation. Ce concept est bien résumé dans le slogan pour les chips Lay's: «Je parie que vous ne pouvez pas en manger juste une!» ("Bet you can't eat just one!"). Le rôle possible des sodas et autres boissons sucrées dans le développement de l'obésité est une question d'actualité. Une méta-analyse de 88 études a trouvé une association de la consommation de sodas avec l'augmentation de l'apport énergétique et l'augmentation du poids corporel.<sup>391</sup> La question de taxer ou non ces boissons reste en débat dans la mesure où une partie des obèses consomme déjà des sodas avec de l'aspartame et qu'il est possible de compenser par d'autres aliments riches en calories.<sup>392,393</sup>



La politique, la religion, l'industrie et la médecine influencent les croyances populaires et les choix personnels tels que nourrir son enfant au sein ou au biberon. L'impact sur le taux d'allaitement est réel avec le taux le plus bas de l'histoire observé à la fin des années 60 et au début des années 70, soit deux décennies avant l'épidémie d'obésité.<sup>394</sup> On pense que la production de lait maternel devrait aider à la perte de poids post-partum acquis pendant la grossesse. L'allaitement maternel exclusif est estimé à 626 kcal par jour.<sup>395</sup> Par conséquent, l'allaitement devrait être davantage encouragé chez les femmes en surpoids et obèses. Cependant, les femmes obèses sont moins susceptibles d'engager avec succès l'allaitement maternel et la durée de l'allaitement est relativement courte si tel est le cas.<sup>396,397</sup> Un IMC plus élevé chez la mère est associé à un retard de lactogénèse de stade II, le moment où la production de lait devient abondante.<sup>396</sup> Des travaux ultérieurs ont montré que 48 heures après l'accouchement, les femmes ayant un IMC>26 sécrétaient moins de prolactine (une hormone essentielle à la lactation) que celles de plus faible IMC. Plusieurs études suggèrent aussi que l'allaitement réduit le risque de développement de l'obésité chez l'enfant. Une méta-analyse de 17 études a trouvé une relation entre la durée de l'allaitement maternel et la réduction du risque de surpoids avec un odds ratio de 1 pour <1 mois d'allaitement maternel, 0,81 pour 1-3 mois, 0,76 pour 4-7 mois, et de 0,67 et 0,68 pour 7-9 et >9 mois.<sup>398</sup> L'odds ratio pour le surpoids était de 0,96 par mois d'allaitement maternel, indiquant une réduction de 4% du risque pour chaque mois de l'allaitement maternel. Cependant, toutes les études ne montrent pas cet effet. Une étude française récente, ELANCE, souligne la nécessité de prendre en compte l'alimentation durant les deux premières années de la vie.<sup>399</sup> Ainsi, l'allaitement maternel a un effet protecteur sur le risque d'obésité à 20 ans seulement si l'alimentation suivie durant les 2 premières années du nourrisson est adaptée (riche en graisses et pauvre en protéines). Cet effet positif pourrait avoir un lien avec le microbiote intestinal.<sup>400,401</sup>

## 8) Causes génétiques des formes communes

Deux opinions opposées sur les causes de l'obésité sont parues en commentaire dans le British Medical Journal (BMJ).<sup>402,403</sup> Le premier auteur plaide en faveur des causes environnementales en évoquant le fait que la prévalence de l'obésité a augmenté principalement durant les 30 dernières années. Pour lui, un tel changement ne peut être dû qu'à une modification de l'environnement, notamment une diminution du coût de l'alimentation et une diminution de l'activité physique. Le deuxième auteur stipule que quand bien même l'environnement serait obésogène, certains individus restent minces et d'autres sont obèses. Son dernier argument concerne les études de jumeaux qui permettent d'estimer l'héritabilité d'une maladie. L'héritabilité de l'adiposité basée sur ces études est estimée entre 40 et 70%.<sup>404-406</sup> Les études d'adoption montrent aussi une forte corrélation entre l'IMC des enfants adoptés et celui des parents biologiques, mais pas avec les parents adoptifs.<sup>407</sup>

Les études de liaison et de gènes candidats ont identifié un certain nombre de gènes et régions génomiques associés à l'obésité, mais ces études atteignent rapidement leurs limites.<sup>408</sup> Les GWAS ont permis d'avoir une approche sans hypothèses et avec une très bonne résolution. Cette méthode interroge plusieurs milliers de SNPs fréquents et répartis dans le génome humain dans le but d'identifier ceux qui sont plus fréquents chez les obèses par rapport aux contrôles. C'est ainsi que depuis 2007, plusieurs vagues de GWAS et de méta-analyses de GWAS ont permis d'identifier une

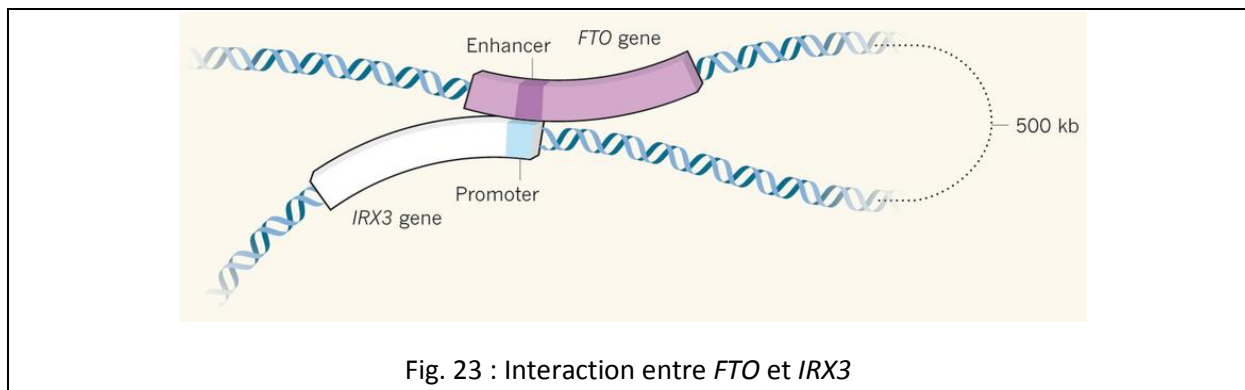
cinquantaine de loci/gènes de susceptibilité à l'obésité. La majorité de ces études portaient sur la variabilité génétique de l'IMC et ont identifié 32 loci associés à l'obésité.<sup>409</sup> Le consortium GIANT a analysé d'autres traits quantitatifs reliés à l'obésité, tels que le tour de taille, le taux de masse grasse ou le ratio taille/hanche pour révéler 13 loci supplémentaires.<sup>410</sup> En stratifiant par le sexe et par les classes d'obésité, 3 loci étaient associés à l'obésité uniquement chez les femmes et 7 autres loci sont spécifiques de certaines classes d'obésité.<sup>411,412</sup> D'un point de vue global, il n'existe, au final, qu'un chevauchement partiel entre les loci/gènes associés aux différents paramètres reliés à l'obésité, ce qui suggère que les voies métaboliques régulant ces phénotypes sont parfois différentes. Une étude portant uniquement sur l'obésité de l'enfant a associé 7 loci précédemment associés à l'obésité de l'adulte et 2 nouveaux loci qui ont un possible rôle dans l'intestin.<sup>413</sup> Si la majorité des loci a été associée à l'obésité et répliquée dans diverses ethnies, 8 loci ont été associés seulement dans des populations d'Asie de l'est<sup>414-416</sup> ce qui laisse présager la découverte de loci spécifiques d'autres ethnies.

Au-delà des associations de SNPs, les puces ont aussi permis d'identifier des Copy-number variations (CNVs). Les CNVs résultent de réarrangements chromosomiques et se traduisent par des délétions, des duplications, des inversions et des translocations. La réanalyse des puces à ADN utilisées dans les GWAS d'obésité a identifié une large délétion de plus de 593kb au niveau du chromosome 16p11.2, augmentant de 43 fois le risque d'obésité, et présent chez 0,7 % des individus obèses morbides européens.<sup>417</sup> Une étude sur l'obésité infantile a mis en évidence 17 CNVs très rares qui comprennent à la fois des délétions et des duplications et qui soulignent le rôle des CNVs dans l'obésité.<sup>418</sup>

Le challenge majeur des GWAS reste l'interprétation biologique des résultats. Lorsque le tagSNP associé à l'obésité est un variant non-synonyme, l'interprétation est relativement aisée. Mais la plupart du temps, comme les SNPs associés sont situés dans des régions non codantes et ont des effets très faibles, il est difficile de savoir s'ils sont bel et bien corrélés au gène le plus proche, et non à un gène situé plus loin sur le chromosome (en cis), voire sur un autre chromosome (en trans). On peut citer un SNP situé dans le premier intron du gène *FTO* et qui a la plus forte association avec l'obésité. On pensait que cette association était liée à une modification de la fonction ou du niveau d'expression de *FTO*, qui code pour une enzyme connue pour être impliquée dans le contrôle du poids et du métabolisme de la souris.<sup>419,420</sup> Toutefois, la preuve d'un lien direct entre ce SNP et l'altération de la fonction *FTO* ou de son niveau d'expression faisait défaut. Un article récent paru dans *Nature* présente des preuves convaincantes que l'association entre l'obésité et *FTO* implique en réalité un autre gène appelé *IRX3*, qui code pour un facteur de transcription impliqué dans les processus de développement multiples.<sup>421</sup> *IRX3* est relativement distant, soit environ 500 kilobases, du premier intron de *FTO*, dans lequel le SNP associé à l'obésité est situé. En utilisant une technique de capture de la conformation du chromosome qui mesure la fréquence d'interactions entre différentes régions du génome, les auteurs ont trouvé que le premier intron de *FTO* interagit physiquement avec le promoteur d'*IRX3* et agit comme un amplificateur (enhancer), régulant ainsi son expression. (Fig. 23)

Enfin, la dernière limite des GWAS repose sur le principe de base, l'utilisation de tagSNPs. En effet, certaines zones du génome sont des déserts de SNPs et les CNVs complexes multi-alléliques sont plus difficiles à détecter. Ainsi, en utilisant des puces transcriptomiques pour étudier le tissu adipeux, une étude récente a mis en évidence une association du gène *AMY1* codant pour l'amylase

salivaire avec l'IMC.<sup>422</sup> L'odd ratio par copie du gène est de 1.19, soit un risque d'obésité augmenté de 19% pour chaque copie en moins du gène. Un grand nombre de copies protège donc de l'obésité et serait dû à l'évolution vers un régime riche en amidons.<sup>423</sup> Cette enzyme catalyse l'hydrolyse de l'amidon dans la salive et aurait aussi un lien avec le microbiote intestinal qui, chez des souris soumises à un régime riche en graisse et en sucre, semble enrichi en *Enterobacteriaceae*, une famille de bactéries associées à l'obésité humaine.<sup>424,425</sup> Cette étude est la première à faire le lien entre le métabolisme des glucides et l'IMC, et expliquerait 2,47–19,86% de la variation génétique de l'obésité.



## 9) Causes génétiques des formes rares

On sait depuis longtemps que l'obésité humaine pouvait résulter du défaut d'un seul gène. Des syndromes cliniquement définis tels que Prader-Willi et Bardet-Biedl étaient fortement associés à l'obésité et un appétit vorace, mais la coexistence de retard de développement et de dysfonctionnement cérébral plus généralisé compliquait l'interprétation. Cependant, au milieu des années 90, les premiers défauts génétiques mis en évidence chez l'homme conduisaient à l'obésité sévère, en l'absence de retard de développement.<sup>426,427</sup> Ces découvertes ont été grandement facilitées par une explosion de la recherche dans le contrôle de l'équilibre énergétique chez les modèles animaux obèses et en particulier par la découverte de la leptine influençant le contrôle central de l'équilibre énergétique, et l'identification du système de la mélanocortine comme un médiateur majeur de l'action de la leptine<sup>303</sup>. Les études familiales et en particulier celles sur les familles consanguines dont les membres souffrent d'obésité extrême et dans la plupart des cas accompagnés d'autres caractéristiques phénotypiques, comme les cheveux roux, l'hyperinsulinisme et l'infertilité, se sont avérées être un succès dans la découverte de gènes de l'obésité humaine.<sup>428,429</sup> On connaît maintenant au moins 34 gènes qui peuvent conduire à une forme autosomique de l'obésité humaine. (**Fig. 24 et 25**) Jusqu'à présent, tous ces défauts affectent le contrôle central de l'équilibre énergétique.<sup>430</sup>

Les formes récessives d'obésité sont causées par la perte de fonction de mutations homozygotes ou hétérozygotes composites dans cinq gènes (*LEP*, *LEPR*, *POMC*, *PCSK1* et *MC4R*). La déficience complète de ces cinq gènes se traduit invariablement par une hyperphagie sévère et des formes pénétrantes d'apparition précoce et extrême d'obésité humaine. D'autres caractéristiques phénotypiques associées à ces formes récessives d'obésité sont plus spécifiques: les individus avec

une déficience en leptine ou son récepteur présentent des altérations de la fonction immunitaire et des infections infantiles fréquentes des voies respiratoires associées à des taux élevés de mort prématurée. Ils présentent également une puberté retardée due à un hypogonadisme central, et une hypothyroïdie.<sup>427,431</sup> Les personnes atteintes de déficience en POMC présentent dès le plus jeune âge une insuffisance surrénale secondaire à un déficit en ACTH et développent une hypoglycémie, un ictère et une mortalité néonatale qui est associée à une choléstase sévère du foie. Les personnes d'ascendance européenne ont la peau pâle et les cheveux roux<sup>432</sup>, mais n'est pas systématique dans d'autres origines ethniques<sup>433</sup>. Les personnes présentant un déficit de PCSK1 ont une hypoglycémie et des perturbations de la fonction endocrinienne notamment des épisodes de diarrhée grave sévère qui indiquent un rôle important de PC1 dans les cellules endocrines.<sup>426,434</sup> Les personnes présentant un déficit complet de MC4R ont une augmentation de la masse maigre, une augmentation de la densité minérale osseuse et une grande taille.<sup>435</sup> Des mutations hétérozygotes dans les gènes *MC4R*<sup>436</sup>, *POMC*<sup>437</sup> et *PCSK1*<sup>438</sup> sont associées à des formes d'obésité à pénétrance incomplète tandis qu'une déficience partielle de LEP ou LEPR est associé à un plus haut pourcentage de masse grasse.<sup>439,440</sup>

Gène	Protéine	Modèle	Caractéristiques cliniques associées	Princeps
<i>BDNF</i>	Brain-derived neurotrophic factor	dominant	retard de développement, hyperactivité, troubles de la mémoire et de la sensation de douleur	441
<i>CEP19</i>	Centrosomal protein of 19 kDa	récessif	DT2, maladies cardiaques, hypertension, stéatose hépatique	442
<i>DYRK1B</i>	Dual specificity tyrosine-phosphorylation-regulated kinase 1B	dominant	coronaropathie à début précoce, hypertension sévère, DT2	443
<i>KSR2</i>	Kinase suppressor of Ras 2	dominant	hyperphagie, rythme cardiaque faible, taux métabolique basal réduit et insulino-résistance sévère	444
<i>LEP</i>	Leptine	récessif	Hypogonadisme, infections fréquentes	427
<i>LEPR</i>	Leptin receptor	récessif	Hypogonadisme, infections des voies respiratoires	431
<i>MC4R</i>	Melanocortin receptor 4	récessif	Accélération de la croissance, augmentation de la taille	445,446
<i>MRAP2</i>	Melanocortin-2 receptor accessory protein 2	dominant		447
<i>NTRK2</i>	BDNF/NT-3 growth factors receptor	dominant	retard de développement, hyperactivité, troubles de la mémoire et de la sensation de douleur	448
<i>PCSK1</i>	Neuroendocrine convertase 1	récessif	Hypoglycémie postprandiale, hypogonadisme	426
<i>POMC</i>	Pro-opiomelanocortin	récessif	Hypopigmentation	432
<i>SH2B1</i>	SH2B adapter protein 1	dominant	résistance à l'insuline, petite taille, troubles du comportement	449
<i>SIM1</i>	Single-minded homolog 1	dominant	Spectre de retard de développement, Prader-Willi like	450,451

Fig. 24 : Causes génétiques des formes rares non syndromiques d'obésité

L'haploinsuffisance de *BDNF*, *NTRK2* et *SIM1* a été associée à une obésité sévère hyperphagique accompagnée de caractéristiques syndromiques chez l'homme.<sup>448,452-454</sup> *BDNF* et son récepteur *TrkB* sont impliqués dans la prolifération, la survie et la différenciation des neurones et permettent de développer la plasticité synaptique dans les neurones hypothalamiques. *SIM1* est un facteur de transcription jouant un rôle important dans la différenciation neuronale du noyau paraventriculaire de l'hypothalamus, une zone du cerveau critique pour la régulation de la prise alimentaire. Aucun cas humain de déficience complète de *BDNF*, *NTRK2* ou *SIM1* n'est rapporté dans la littérature, ce qui est cohérent avec les phénotypes observés chez les souris. Un déficit partiel de *Bdnf*, *Ntrk2* ou *Sim1* chez la souris induit une hyperphagie, une obésité et des caractéristiques de développement,<sup>455-457</sup> alors qu'un déficit complet de *Bdnf* ou *Sim1* est létal<sup>455,456</sup> et un déficit complet de *Ntrk2* réduit considérablement la durée de vie<sup>458</sup>. Chez l'homme, des mutations dans *SIM1* causent une obésité à pénétrance variable et parfois avec un phénotype similaire observé dans le syndrome de Prader-Willi.

Une étude récente montre que des mutations délétères dans *SH2B1* sont associées à une obésité sévère à début précoce, une résistance à l'insuline, une petite taille, et un ensemble de troubles du comportement, y compris une tendance à l'isolement social et des niveaux élevés d'agression.<sup>449</sup> Une des mutations retrouvées (une mutation frameshift) modifie la signalisation de la leptine, mais les autres mutations n'ont pas cet effet. Cette discordance peut refléter une mauvaise sensibilité des tests utilisés, ou certains des effets de *SH2B1* peuvent être dus à une voie de signalisation indépendante de la leptine.

Des mutations du gène *MRAP2* sont associées à une obésité sévère de l'enfant. Les études sur la souris et le poisson-zèbre laissent à penser que *Mrp2* peut modifier la signalisation par *MC4R* et exercer ainsi des effets sur la régulation du poids.<sup>447,459</sup>

Le gène *KSR2* ne semble pas être directement lié au système leptine-mélanocortine. Les porteurs de mutation délétère dans ce gène présentent une hyperphagie dès l'enfance, un rythme cardiaque faible, un taux métabolique basal réduit et une insulino-résistance sévère. Les variants rares dans *KSR2* perturbent la signalisation de la voie Raf-MEK-ERK et nuisent à l'oxydation des acides gras cellulaires et de l'oxydation du glucose. Ces effets peuvent être améliorés par la prise d'un antidiabétique, la metformine. Les variants retrouvés ne coségrégent pas dans les familles obèses et cette pénétrance incomplète suggère que d'autres facteurs génétiques et environnementaux jouent également un rôle.

Dans la même lignée, les mutations délétères du gène *DYRK1B* semblent avoir un effet pléiotropique : obésité abdominale, coronaropathie à début précoce, hypertension sévère, DT2. De la même façon que *KSR2*, *DYRK1B* inhibe la voie RAS-RAF-MEK qui régule l'absorption du glucose et la glycolyse. De plus, *DYRK1B* pourrait augmenter l'adipogenèse par l'inhibition de la voie SHH ou par l'augmentation du renouvellement de p27Kip, ce qui pourrait expliquer l'athérosclérose observée également chez les souris. En résumé, les auteurs pensent avoir un gène dont la perturbation conduit au syndrome métabolique. La contribution de ce gène pourrait être plus large dans la population générale sachant que le locus a été associé à la fois au DT2 et syndrome métabolique dans les GWAS.

Enfin, un dernier gène, *CEP19*, ne semble également pas lié au système leptine-mélanocortine et pourtant son inactivation conduit à la fois à une obésité sévère chez l'homme et chez la souris. Les mécanismes moléculaires de la protéine ciliaire *CEP19* ne sont pas connus mais

son rôle semble crucial car il existe une grande conservation de sa séquence protéique au cours de l'évolution. Le phénotype humain observé va au-delà de l'obésité morbide puisque les porteurs développent un syndrome métabolique (DT2, dyslipidémie, hypertension, maladies cardiaques telles que la coronaropathie et l'infarctus du myocarde). Trois membres atteints présentent une déficience intellectuelle et tous les hommes atteints avaient une diminution du nombre de spermatozoïdes.

Les mutations décrites précédemment sont toutes rares et conduisent, en plus de l'obésité à des manifestations phénotypiques supplémentaires. Ces effets pléiotropiques se rapprochent des formes syndromiques d'obésité, mais le phénotype majeur reste l'obésité, et les phénotypes supplémentaires sont très variables et n'impliquent pas (ou peu) un retard de développement comme dans ces formes syndromiques. Cinq formes syndromiques sont bien connues : Alström, Bardet-Biedl, McCune-Albright, Cohen et Prader-Willi. La sixième forme est moins connue, car très rare et décrite une seule fois dans la littérature.

Gène	Syndrome	Modèle	Caractéristiques cliniques associées	Princeps
<i>ALMS1</i>	Alström	récessif	Photophobie, nystagmus, troubles visuels, surdité, insulino-résistance sévère	460,461
<i>BBS1-18</i>	Bardet-Biedl	récessif	Polydactylie, dystrophie rétinienne, hypogonadisme, anomalies rénales, retard de développement	462
<i>GNAS</i>	McCune-Albright	dominant	Petite taille, anomalies du squelette	463
<i>TUB</i>		récessif	cécité nocturne, baisse de l'acuité visuelle, dystrophie mixte	464
<i>VPS13B (COH1)</i>	Cohen	récessif	retard mental, microcéphalie, myopie	465
15q11–q13	Prader-Willi	dominant	petite taille, petites mains et petits pieds, retard mental, hypogonadisme, et hyperphagie	466

Fig. 25 : Causes génétiques des formes rares syndromiques d'obésité

Le syndrome de Prader-Willi (PWS) est peut-être le plus commun des syndromes d'obésité (prévalence estimée à 1/25 000 naissances). C'est une maladie caractérisée par une obésité, une activité fœtale réduite, une hypotonie musculaire à la naissance, une petite taille, des petites mains et petits pieds, un retard mental, un hypogonadisme, et une hyperphagie. Cette maladie est causée soit par une délétion du segment chromosomique 15q11-q13 paternel, soit par la perte de l'ensemble du chromosome paternel 15. Plusieurs gènes candidats dans la région délétée ont été étudiés, mais une étude récente indique que le gène *MAGEL2* serait responsable en partie du phénotype de Prader-Willi dont l'obésité.<sup>467</sup> Cependant, une autre étude souligne le rôle d'un ARN non codant (lncRNA) appelé *HOTAIR* qui, s'il est surexprimé comme dans le syndrome de Prader-Willi, régule la région *DLK1-DIO3* située sur le chromosome 14 et conduit à une surexpression de 6 gènes de la famille des MEGs (maternally expressed genes).<sup>468</sup>

Le syndrome de Bardet-Biedl (BBS) est un syndrome rare (prévalence estimée à <1/100 000 naissances) caractérisé par une obésité, un retard mental, une rétinopathie pigmentaire, une

polydactylie, un hypogonadisme et des anomalies rénales. Beaucoup de ces gènes semblent affecter les protéines localisées sur le corps basal, un élément clé des cils primaires (prolongements cellulaires), jugées importantes pour l'interaction intercellulaire dans les cellules de mammifères, y compris les neurones.<sup>469</sup> Une étude a suggéré un lien entre la fonction ciliaire et la signalisation de la leptine.<sup>470</sup>

Le syndrome de McCune-Albright (AHO) est une maladie caractérisée par une obésité, une petite taille, un visage rond, une brachydactylie, une formation osseuse dans les tissus non osseux, un retard mental, et une hypocalcémie. Sa prévalence est de 1/139 000 au Japon. L'AHO résulte de mutations germinales dans *GNAS1* qui diminuent l'expression ou la fonction de la protéine Gs $\alpha$ . Les souris dépourvues de Gs $\alpha$  sont obèses, et de façon similaire aux souris déficientes en MC4R, elles ont une réduction du tonus sympathique, mais ne sont pas hyperphagiques.<sup>471</sup>

Le syndrome de Cohen est dû à une déficience d'un gène qui a été localisé sur le chromosome 8q dans plusieurs familles finlandaises. Ce syndrome est caractérisé par une obésité, un retard mental, une microcéphalie, des traits caractéristiques du visage, et une myopie. Des mutations du gène *COH1* conduisent au syndrome Cohen dans différents groupes ethniques.<sup>472</sup>

Le syndrome d'Alström est une maladie de l'enfance caractérisée par une obésité, une dystrophie rétinienne, une cardiomyopathie dilatée, un diabète de type 2 et d'autres traits cliniques de gravité variable, tels qu'une hypothyroïdie, une petite taille, un hypogonadisme, des anomalies rénales, et pour certains un retard mental. Des mutations dans le gène *ALSM1* sont responsables de ce syndrome.

Les rétinites pigmentaires ou *retinitis pigmentosa* sont des maladies de l'œil qui sont parfois liées à des syndromes en combinaison avec l'obésité comme les syndromes d'Alström et Bardet-Biedl. Une étude récente décrit une nouvelle ciliopathie caractérisée par une dystrophie rétinienne et une obésité précoce due à une mutation homozygote dans *TUB*.<sup>464</sup> Ce gène semble jouer un rôle dans la signalisation des récepteurs couplés aux protéines G (RCPG) au niveau des cils et dans la modulation de la signalisation par les RCPG hypothalamiques impliqués dans l'homéostasie énergétique et le métabolisme du glucose par une modulation des effets de l'insuline et de la leptine.<sup>473</sup>



## C. Génétique des maladies rares

### 1) Analyse de liaison

La liaison et le déséquilibre de liaison sont deux concepts étroitement liés. Deux loci génétiques sont liés s'ils sont transmis ensemble des parents à l'enfant plus souvent que lors d'une ségrégation indépendante, c'est-à-dire avec une probabilité de recombinaison inférieure à 50%. Les loci sont en déséquilibre de liaison si dans une population générale, ils sont trouvés ensemble sur le même haplotype plus souvent qu'attendu. En général, deux loci en déséquilibre de liaison sont aussi liés mais l'inverse n'est pas nécessairement vrai.<sup>474</sup>

L'analyse de liaison est souvent la première étape dans l'investigation génétique d'un trait, car elle peut être utilisée pour identifier des grandes régions génomiques qui peuvent contenir un gène d'intérêt associé à la maladie. C'est à la fois une méthode biologique et statistique qui permet une approche sans hypothèse sur l'ensemble du génome. Cette méthode repose sur l'utilisation de marqueurs biologiques, par exemple les STRs (short tandem repeats) ou microsatellites. Les STRs sont des motifs répétés souvent de 2 à 4 pb qui constituent une séquence ADN souvent inférieure à 100 pb qui peut être facilement amplifiée par PCR pour évaluer sa taille à partir d'une quantité minimale d'ADN génomique. Ce sont des marqueurs de choix, car ils sont très polymorphes et répartis sur l'ensemble du génome. Toutefois, le génotypage des STRs est difficilement automatisable et les SNPs, bien que moins polymorphes, mais plus faciles à génotyper, ont remplacé les STRs notamment avec l'utilisation des méthodes TaqMan et SNPplex.

Les études basées sur des familles pour étudier un effet génétique ont plus de puissance statistique que dans une étude cas/contrôle avec le même nombre d'individus.<sup>475</sup> Cependant, le recrutement de familles est plus difficile que celui d'une population générale. Plusieurs types d'étude familiale sont possibles : les trios parents-enfants, des fratries (concordantes ou discordantes pour un trait), des familles étendues avec plusieurs membres atteints, ou encore des familles consanguines. Les études de liaison reposent sur une évaluation statistique de la coségrégation du marqueur biologique avec un trait dans une étude familiale. En fonction du design de l'étude, de la question posée ou des objectifs de l'étude, plusieurs techniques d'analyse sont possibles.

L'analyse de liaison paramétrique suppose un modèle particulier de la maladie et utilise des méthodes de vraisemblance pour tester la ségrégation non indépendante des marqueurs et de la maladie.<sup>476</sup> On peut prendre en compte plusieurs paramètres : un modèle récessif ou dominant, l'âge de début de la maladie, la fréquence allélique, ou encore l'hétérogénéité de la maladie par la pénétrance en intégrant un certain pourcentage de phénocopies. Cette analyse permet le calcul d'un score LOD (logarithme des probabilités) qui est une fonction du taux de recombinaison ou la position chromosomique mesurée en cM.<sup>477</sup> Plus le score est élevé et positif, plus la liaison est probable.

L'analyse de liaison non paramétrique ne nécessite pas de connaître le modèle de la maladie. Elle compare l'identité par descendance (IBD, si les 2 allèles sont issus d'un ancêtre commun sans recombinaison) entre les membres atteints à un locus spécifique du génome, estimée à partir des données des marqueurs, pour voir si elles sont significativement différentes des prédictions de Mendel.<sup>476</sup> Une ségrégation non aléatoire des marqueurs et du trait est attendue si les marqueurs sont liés à la maladie. La cartographie par homozygotie est une autre technique qui est souvent considérée comme une approche non paramétrique de l'analyse de liaison. Cette approche suppose

un allèle rare responsable de la maladie et un mode récessif. Parce que cette approche s'appuie simplement sur la détection de grandes zones d'homozygotie chez les individus atteints, elle a l'avantage de ne nécessiter aucune connaissance exacte des familles, par exemple l'apparition de la consanguinité, pour l'analyse de liaison.

Les expériences passées ont montré que cette approche est très fiable. Le principal avantage de l'analyse de liaison est l'identification d'une petite région candidate ce qui augmente la probabilité a priori qu'un variant identifié dans cette région soit causal. Le principal inconvénient de l'analyse de liaison est l'obligation d'étudier de grandes familles. Pour atteindre une signification statistique, une famille multigénérationnelle avec 6-12 personnes atteintes est nécessaire pour un modèle autosomique dominant. Plus le nombre de sujets est important, plus la région identifiée sera petite ce qui facilitera les analyses ultérieures. Les maladies autosomiques récessives dans des familles consanguines peuvent souvent être résolues dans de petites familles à l'aide de la cartographie par homozygotie. Toutefois, cette analyse reste limitée dans le cas de mutations hétérozygotes composites, car les enfants atteints n'héritent pas de copies identiques des allèles transmis par les parents. En outre, des phénocopies et une faible pénétrance de la maladie empêchent fortement l'identification des gènes par l'analyse de liaison.

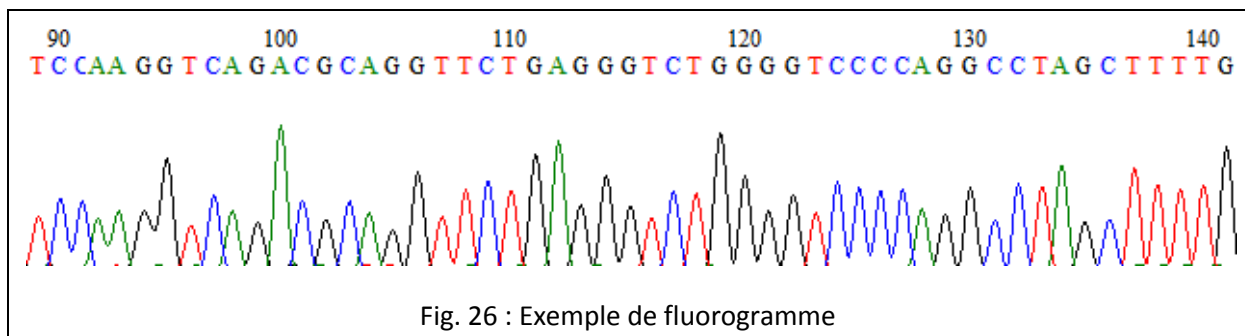
L'analyse de liaison permet seulement l'identification d'une région génomique qui peut contenir près d'une centaine de gènes. L'étape suivante est généralement de rajouter des marqueurs pour restreindre la zone d'intérêt (fine mapping). On peut ensuite coupler cette technique au clonage positionnel de la région d'intérêt dans un BAC (Bacterial Artificial Chromosome) et à « l'exon trapping » pour déterminer les fragments qui correspondent à un gène exprimé. Les données physiologiques et moléculaires peuvent aussi aider à sélectionner un gène candidat. La seule façon d'augmenter la résolution et d'identifier une mutation est le séquençage. Si le séquençage Sanger est couramment utilisé, on peut aussi utiliser les résultats de l'analyse de liaison pour restreindre le nombre de mutations obtenues par séquençage nouvelle génération (NGS). On peut par exemple citer la découverte du gène *GUCY2C* mis en jeu dans une forme dominante de diarrhée familiale<sup>478</sup> ou encore le gène *DYRK1B* dont la perturbation conduit au syndrome métabolique<sup>443</sup>. De récents développements algorithmiques témoignent que l'analyse de liaison n'est pas encore abandonnée en l'état et présente même un intérêt pour du séquençage de génome entier.<sup>479</sup>

## 2) Séquençage Sanger

En 1977, Fred Sanger et Alan R. Coulson ont publié deux papiers méthodologiques sur la détermination rapide d'une séquence d'ADN<sup>480,481</sup>, ce qui transformera la biologie dans son ensemble en fournissant un outil pour déchiffrer des gènes complets et plus tard des génomes entiers. La méthode a considérablement amélioré les techniques de séquençage d'ADN précédemment développées par Maxam et Gilbert et publiée la même année<sup>482</sup>, et celle de Sanger et Coulson publiée deux années plus tôt<sup>483</sup>. Les avantages évidents de la réduction de produits chimiques toxiques et de radio-isotopes ont rapidement fait du séquençage Sanger la seule méthode de séquençage d'ADN utilisée dans les prochaines décennies.

La commercialisation du premier séquenceur automatique de la société Applied Biosystems date de 1986. Cette avancée a permis une démocratisation de cette technologie et de concrétiser

plusieurs projets dont le séquençage du génome humain qui a duré 15 ans. Ce séquençage exploite la technique du "Dye terminator" qui est une adaptation de la technique de Sanger et est basée sur l'utilisation de fluorescence. À chaque type de didésoxynucléotide est associé un fluorophore spécifique : rouge pour didésoxyT, noir pour didésoxyG, bleu pour didésoxyC, vert pour didésoxyA. La réaction de séquençage produite par PCR génère des fragments porteurs de fluorophores terminaux. L'AmpliTaQ FS incorpore des dNTPs normaux et des ddNTPs fluorescents ou BigDye Terminators. Lorsqu'un fluorophore est incorporé, la polymérase s'arrête. Comme la concentration des fluorophores diminue au fur et à mesure de la réaction, les fragments générés seront statistiquement de plus en plus longs et de toutes les tailles. Après filtration sur gel de la réaction de séquence et injection électrocinétique des fragments obtenus, ces derniers sont soumis à une électrophorèse capillaire. La migration des fragments fluorescents au travers d'une cellule de détection permet au système optique laser du séquenceur de collecter la fluorescence portée par chacun de ces fragments et de transmettre ces informations à l'ordinateur. Un fluorogramme et un fichier texte sont ensuite générés par correspondance entre les courbes de fluorescence et le nucléotide incorporé. (**Fig. 26**) Un logiciel traduit ces courbes en séquence d'ADN, tout en générant des probabilités d'erreur pour chaque appel de base.



### 3) Séquençage nouvelle génération

Avec l'objectif ultime de déchiffrer le génome humain, le besoin d'augmentation des débits de séquençage a pris une ampleur imprévue, conduite par des développements tels que l'électrophorèse capillaire automatisée. Les automates de laboratoire et les processus de parallélisation ont abouti à la création d'entreprises ressemblant plus à des usines telles que les centres de séquençage qui abritent des centaines d'instruments de séquençage exploités par des cohortes de personnel. Cependant, même la réussite des deux projets de séquençage du génome humain n'a pas satisfait la soif des biologistes pour un séquençage encore plus haut débit et, surtout, une technologie de séquençage plus économique.

Les premiers signes de ce qui pourrait révolutionner le marché de séquençage sont apparus en 2005 avec la publication de la technologie de séquençage par synthèse développée par 454 Life Sciences et le protocole de séquençage multiplexe de polonies (regroupement ponctuel d'ADN) du laboratoire de George Church. Les deux groupes ont utilisé une stratégie qui permet de réduire considérablement le volume de réaction nécessaire tout en augmentant considérablement le nombre de réactions de séquençage. Cette stratégie entraîne un réseau de plusieurs centaines de milliers de matrices de séquençage, dans un support permettant que ces séquences soient analysées en parallèle.

Tous les séquenceurs de nouvelle génération nécessitent une étape de préparation de librairie avant le séquençage qui diffère selon l'application. Dans notre cas, cette étape permet un enrichissement de régions spécifiques qui peuvent se limiter à quelques gènes ou couvrir l'exome humain. Cette étape est réalisée par deux méthodes différentes : l'enrichissement par hybridation ou par PCR.

#### a) Enrichissement par hybridation

Cette méthode compte plusieurs étapes principales de préparation (**Fig. 27**) :

- L'ADN génomique est coupé en petits fragments ; plusieurs méthodes existent, mais la plus couramment utilisée est la sonication basée sur l'utilisation d'ondes sonores
- Des adaptateurs sont liés aux fragments d'ADN soniqué et permettront l'hybridation sur la flow cell durant l'étape de séquençage ; les adaptateurs sont généralement constitués d'un index unique qui permet d'identifier chaque échantillon après séquençage
- Le mix précédemment obtenu est hybridé à des amorces d'ADN ou d'ARN qui correspondent aux zones ciblées d'intérêt et qui sont marquées
- Les fragments d'ADN non hybridés sont lavés
- Les fragments restants sont purifiés et quantifiés

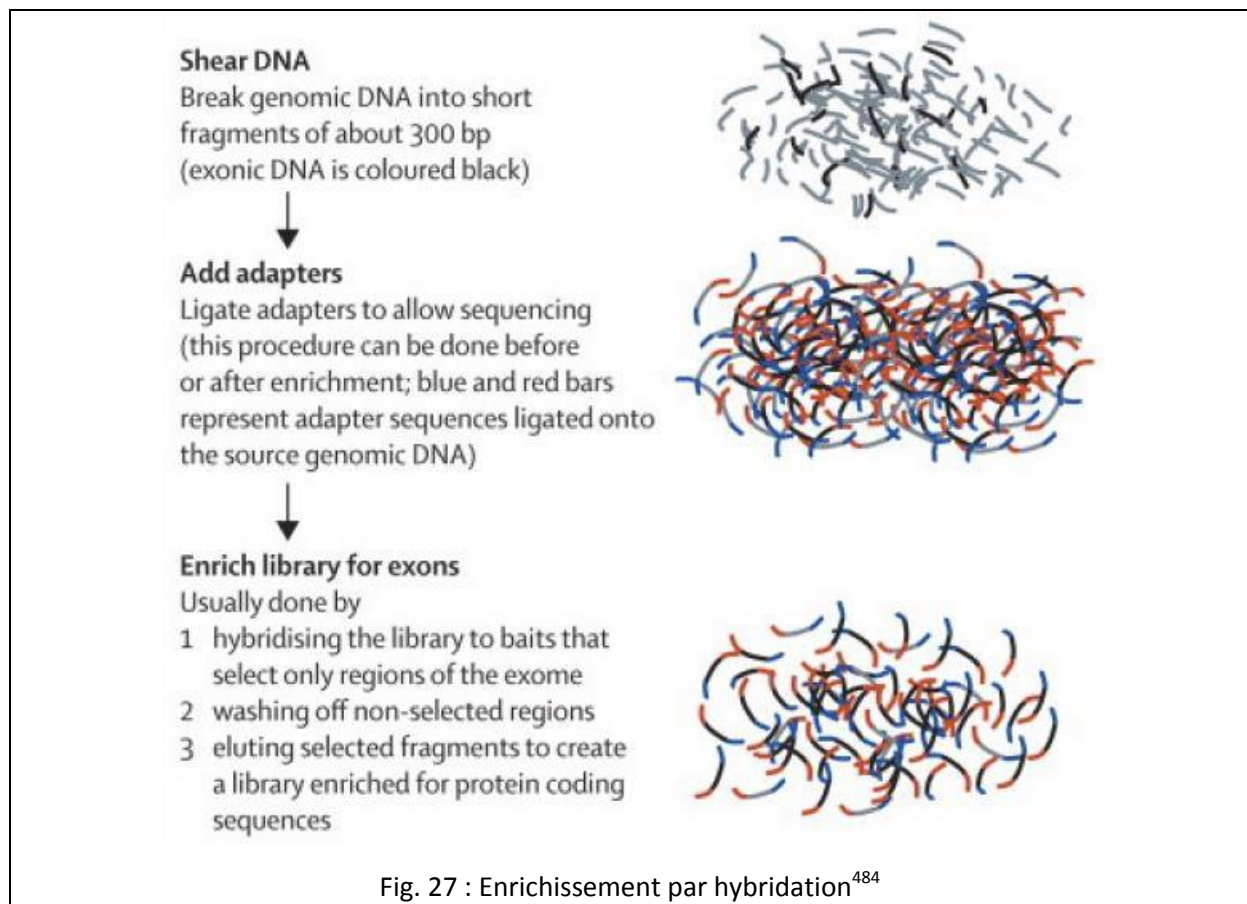


Fig. 27 : Enrichissement par hybridation<sup>484</sup>

Les amorces sont souvent marquées par la biotine qui a une forte affinité pour la streptavidine. Des billes magnétiques recouvertes de streptavidine permettent l'isolation des

fragments d'intérêt. Il existe plusieurs kits commerciaux de capture par hybridation. Pour cibler l'exome, il existe par exemple SureSelect et Haloplex (Agilent), TruSeq (Illumina) et SeqCap EZ (Roche NimbleGen). Les régions ciblées peuvent différer entre les kits par exemple les UTR 5' et 3'. La différence majeure se situe au niveau de la taille des amorces et leur nature (ARN pour SureSelect et ADN pour les autres).

#### b) Enrichissement par PCR

RainDance Technologies, Fluidigm et WaferGen sont les trois entreprises principales qui proposent des techniques d'enrichissement par PCR. Les technologies de Fluidigm et WaferGen sont basées sur l'utilisation de puce microfluidique contenant des milliers de chambres réactionnels alors que celle de RainDance permet l'utilisation de microgouttelettes lipidiques pour isoler différentes réactions chimiques comme la PCR. RainDance a été fondée en 2004 par Darren R. Link, docteur en physique de l'Université Harvard, Andrew Griffiths, docteur en chimie du Centre de recherche médicale de Cambridge, et Jérôme Bibette, docteur en physique/chimie de l'École Supérieure de Physique et Chimie Industrielles de Paris. RainDance développe des outils permettant l'application de gouttelettes microscopiques en microfluidique pour la recherche médicale. Les outils élaborés par RainDance servent essentiellement à la détection de mutations cancérigènes et de CNVs par PCR digital et au reséquençage ciblé du génome humain. Cette dernière application permet l'analyse à haute résolution des variations génétiques entre individus et populations. RainDance développe d'autres applications type « single cell » pour l'isolation de cellules humaines<sup>485</sup>, mais aussi pour des microorganismes, des systèmes pour le criblage à haut débit de bibliothèque de petites molécules et de protéines, et des projets plus fondamentaux tels que l'utilisation des microgouttelettes comme analogues des protocellules pour étudier le rôle de compartimentalisation dans la chimie prébiotique et l'origine de la vie, dans la conservation et le transfert d'information par les ARNs et dans l'origine des chromosomes (transcription et traduction).

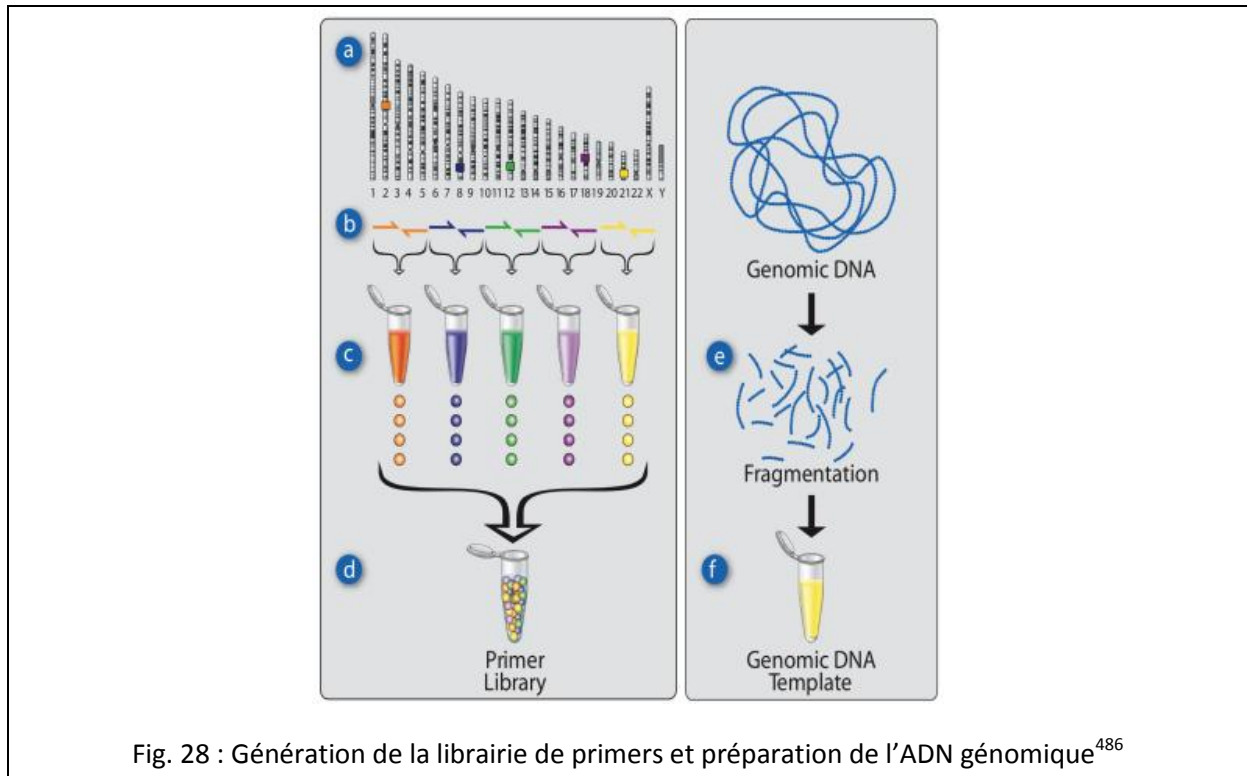
Concernant le séquençage ciblé, le flux de travail (**Fig. 28**) commence par la génération de la librairie et comporte plusieurs étapes :

- (a) Identifier les séquences d'intérêt du génome qu'on veut cibler : on peut fournir une liste de gènes ou même des coordonnées génomiques précises
- (b) Le design des primers est basé sur la plateforme Primer3 avec des critères tels qu'un pourcentage en GC entre 25 et 80%, une taille d'amplicon adaptée au type de séquenceur, un Tm compris entre 56 et 60°C, une taille de primer entre 15 et 33pb, et l'absence de SNP connu dans les primers ; les primers sont ensuite synthétisés par une autre société
- (c) Une puce microfluidique est utilisée pour encapsuler séparément chaque couple de primers dans des microgouttelettes lipidiques de 25 microns ; l'enveloppe lipidique est constituée d'une huile de support inerte fluorée couplée à un copolymère tensioactif
- (d) Les microgouttelettes lipidiques sont mélangées de façon équimolaire pour que chaque couple soit représenté proportionnellement

La deuxième étape constitue la préparation du mix PCR contenant l'ADN génomique :

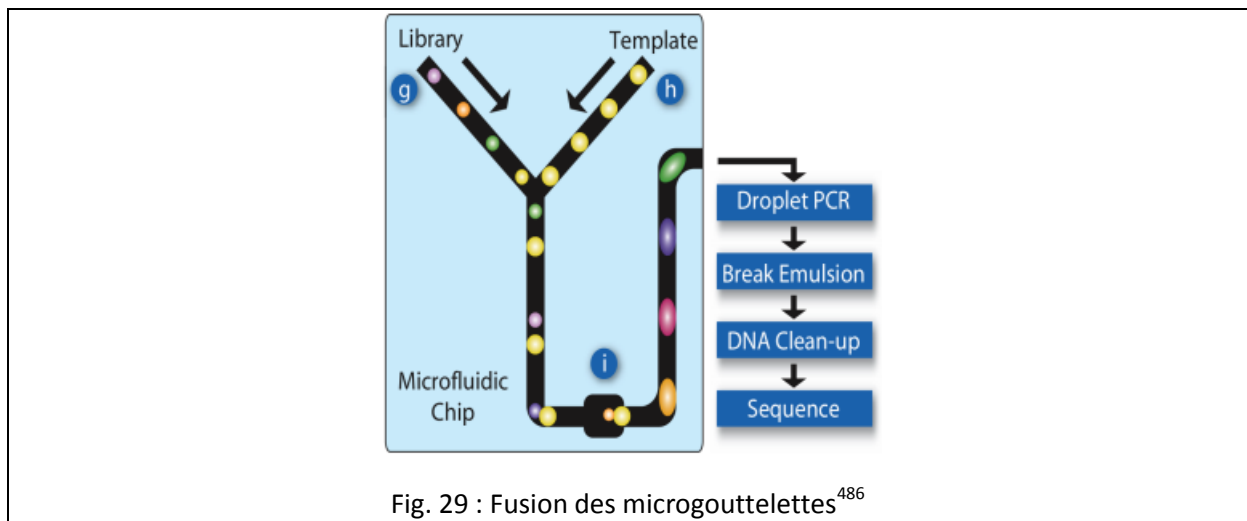
- (e) L'ADN génomique est fragmenté, généralement par sonication, entre 2 et 4 kb puis purifié

- (f) L'ADN fragmenté est mélangé à tous les constituants de la PCR à l'exception des primers :  
buffer, MgSO<sub>4</sub>, mix de dNTPs, bêtaïne, stabilisateur, DMSO, Taq haute fidélité, eau



La dernière étape (**Fig. 29**) à fusionner les microgouttelettes :

- (g) Les microgouttelettes contenant les primers (environ 25 microns) sont injectées dans le circuit de la puce microfluidique
- (h) L'ADN génomique mélangé aux réactifs PCR est injecté dans le circuit de la puce microfluidique pour former des microgouttelettes d'environ 33 microns ; ces microgouttelettes sont ensuite couplées aux microgouttelettes contenant les primers dans un ratio 1 :1
- (i) Un champ électrique permet de fusionner les microgouttelettes (37 microns) ; environ 1 million de microgouttelettes sont produites et récupérées dans un support

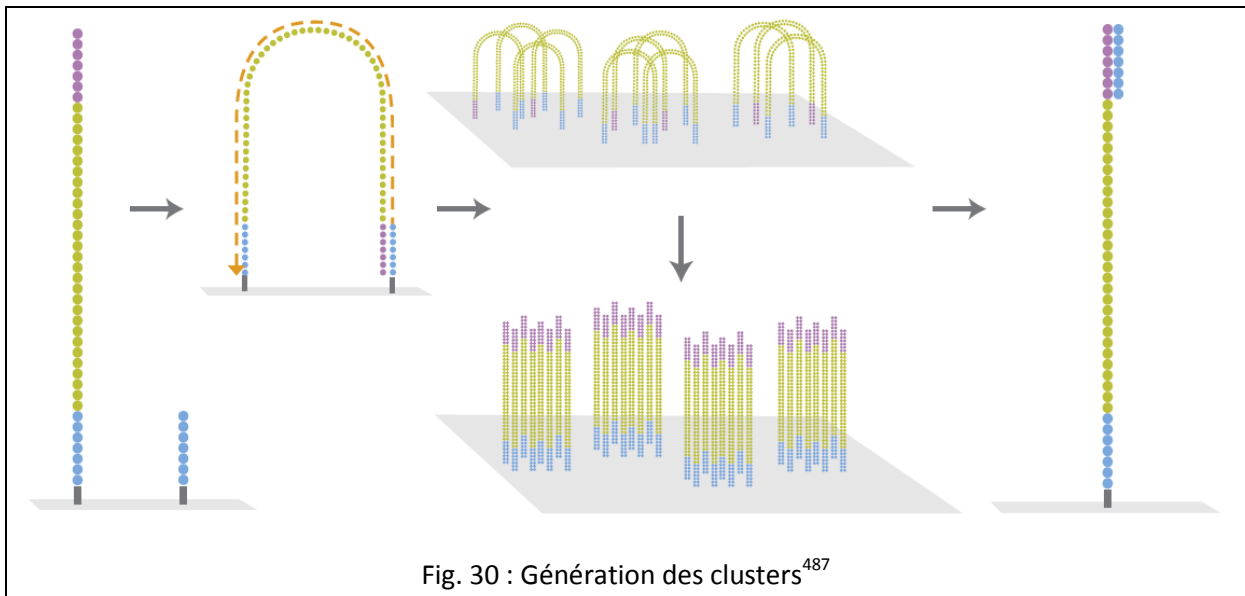




On procède ensuite à une PCR sur un thermocycleur standard pour permettre l'amplification ciblée. Une solution chimique permet de casser l'émulsion et purifier la réaction de séquence.

### c) Séquençage Illumina

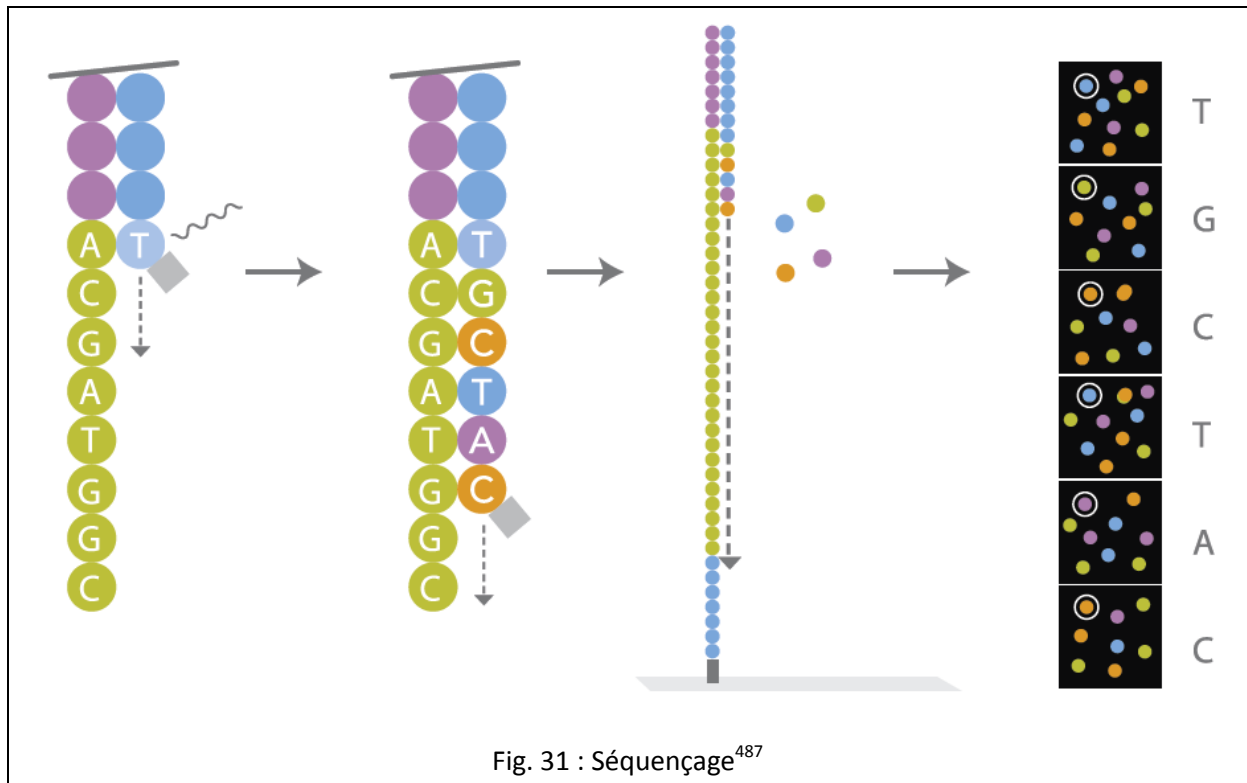
La plateforme de séquençage Solexa a été commercialisée en 2006, et rachetée par Illumina début 2007. Le principe (**Fig. 30 et 31**) est basé sur une chimie de séquençage par synthèse, avec des nucléotides terminateurs réversibles pour les quatre bases d'ADN chacune marquée avec un fluorophore différent, et une ADN polymérase capable de les incorporer.



Après fabrication de la librairie, les fragments d'ADN sont dénaturés et une des 2 extrémités est liée à un support solide. La surface du support est recouverte par une forte densité d'adaptateurs et d'adaptateurs complémentaires. Chaque fragment simple brin, immobilisé par une extrémité sur la surface, crée une structure "en pont" par hybridation avec son extrémité libre à l'adaptateur complémentaire sur la surface du support. Après avoir ajouté un mélange contenant les réactifs d'amplification par PCR, les adaptateurs liés à la surface servent d'amorce pour la PCR. L'amplification est nécessaire pour obtenir suffisamment d'intensité du signal lumineux pour la détection fiable des bases ajoutées. Après plusieurs cycles de PCR, des clusters aléatoires d'environ 1000 copies simples brins d'ADN (aussi appelés colonies) sont créés sur la surface.

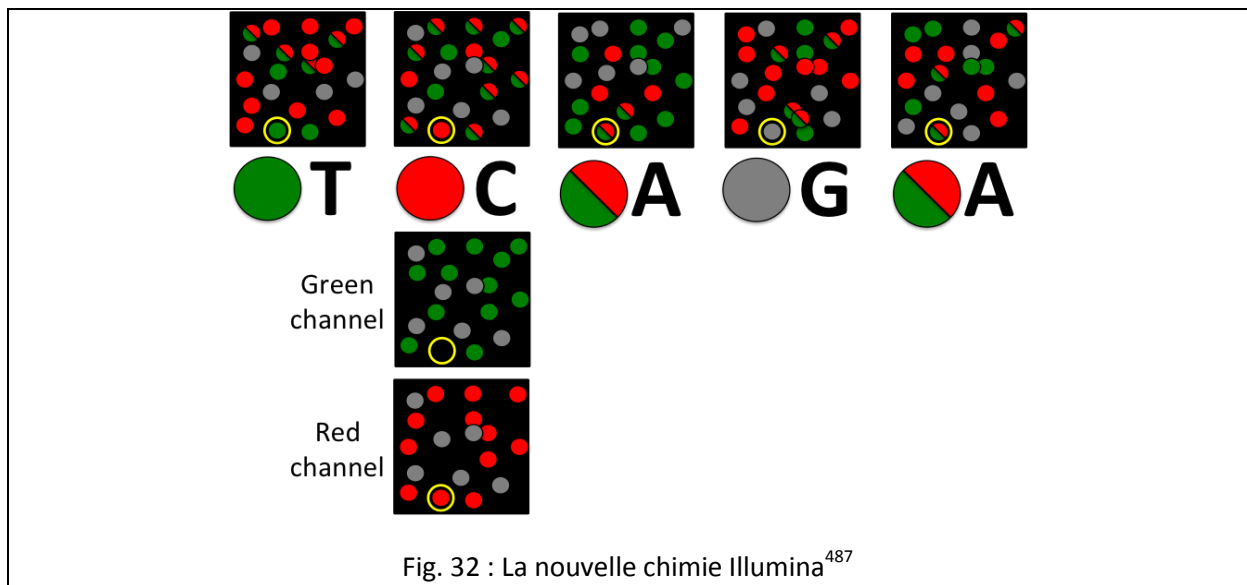
Le mélange réactionnel pour le séquençage et la synthèse d'ADN est distribué sur la surface et contient des amorces, des quatre nucléotides terminateurs réversibles marqués chacun avec un colorant fluorescent différent, et l'ADN polymérase. Après l'incorporation dans le brin d'ADN, le nucléotide terminateur, ainsi que sa position sur la surface de support, est détecté et identifié par l'intermédiaire de son fluorophore identifié par la caméra CCD. Le groupe terminateur à l'extrémité 3' de la base et le fluorophore sont ensuite retirés de la base nucléotidique et le cycle de synthèse est répété. La longueur de la séquence lue était d'environ 35 nucléotides dans la première génération de séquenceur et elle atteint maintenant 300 paires de bases. La séquence de millions de clusters peut être déterminée simultanément en parallèle, ce qui entraîne un débit très élevé de lecture, de l'ordre de Gigabases voire Terabases par support.





Les séquenceurs Illumina ont vu de nombreuses améliorations au fur et à mesure des générations :

- La simplification des protocoles d'utilisation
- L'augmentation de la longueur des séquences lues
- L'augmentation du débit (le nombre de clusters)
- La diminution du temps de run
- L'amélioration de la fiabilité de la chimie (pour les problèmes d'homopolymères, etc.)



Récemment, deux nouveaux séquenceurs ont été introduits sur le marché : le NextSeq 500 et l'HiSeq X Ten. Une modification de la chimie de synthèse (**Fig. 32**) répond simultanément à plusieurs

problématiques, à savoir diminuer par 2 la complexité de production des réactifs, d'instrumentation et de calcul informatique. Les 4 bases nucléotidiques ne sont plus marquées par 4 fluorophores différents, seulement un fluorophore rouge et un fluorophore vert sont utilisés. Ainsi, la thymine est marquée par un fluorophore vert, la cytosine par un fluorophore rouge, l'adénine à la fois par un fluorophore vert et rouge, et la guanine n'est pas marquée du tout. Cette optimisation permet de prendre 2 images au lieu de 4 et diminue par 2 l'espace informatique requis pour le stockage des données. Aujourd'hui, Illumina est un acteur majeur dans le domaine du séquençage et revendique l'utilisation de sa technologie dans plus de 90% des publications scientifiques sur le sujet.

#### d) L'analyse bioinformatique

Ce qui prenait avant des années et de nombreuses collaborations pour générer la séquence du génome humain, peut maintenant être fait en quelques jours par une seule personne en utilisant une des dernières machines de séquençage de dernière génération. L'inconvénient majeur est que cette quantité massive de données se présente sous la forme de courtes séquences (les reads), et il faut s'investir massivement dans l'analyse informatique et choisir parmi une pléthore d'outils pour donner un sens à tout ça. L'analyse des données peut être décomposée en 5 grandes étapes (**Fig. 33**) : le contrôle de la qualité des données brutes, l'alignement des reads sur un génome de référence, l'identification des variants, l'annotation des variants, et la visualisation des données. Ces étapes constituent un pipeline qu'on peut définir comme un outil permettant d'exécuter un ensemble de processus de façon automatique. En amont de ce pipeline, par exemple dans le cas d'Illumina, le logiciel du séquenceur convertit les images obtenues en reads et c'est ce qu'on appelle le base calling. Le logiciel fournit aussi des métriques en temps réel qui permettent d'avoir une première idée de la qualité des données. On a par exemple la densité de clusters, le score Q et le taux d'erreurs. Le score Q est un indice inspiré du score Phred calculé pour le séquençage Sanger qui permet de mesurer la précision et qui correspond dans le cas d'un Q30 à la probabilité qu'une base soit incorrectement appelée 1 fois sur 1000. Le taux d'erreur est calculé sur le génome du PhiX qui est utilisé comme contrôle dans chaque run de séquençage.

La première étape de l'analyse est d'évaluer la qualité des reads bruts et de supprimer, couper (trimming) et corriger les reads qui ne respectent pas les normes définies. Les données brutes générées par les plateformes de séquençage sont compromises par des artefacts de séquence de base tels qu'une mauvaise qualité des reads ou encore une contamination des adaptateurs.<sup>488</sup> Ces erreurs sont très fréquentes dans les données de séquençage, et les plateformes sont sensibles à un large éventail de défaillances de la chimie et des instruments.<sup>489,490</sup> Il est donc nécessaire d'effectuer un filtrage et un trimming pour éviter de tirer des conclusions erronées. En général, ces mesures comprennent la visualisation des scores de qualité par base et la distribution des nucléotides, et d'autres paramètres tels que les contaminations de primers, la teneur en N (nucléotide non connu) et les biais de GC. L'étape de trimming consiste à enlever les séquences des adaptateurs et inclut donc un démultiplexage qui permet grâce aux index utilisés d'apparier les reads à un échantillon. Le format des fichiers de sortie le plus couramment utilisé est le fastq.

Après avoir traité les reads pour répondre à un certain niveau de qualité, ils sont généralement alignés sur un génome de référence existant.<sup>491</sup> Actuellement, il existe deux sources principales de référence du génome humain: l'Université de Santa Cruz (UCSC), qui accueille

également les données ENCODE<sup>492</sup> et le Genome Reference Consortium (GRC), qui se concentre sur l'assemblage de génomes de référence de diverses espèces<sup>493</sup>. Les deux ressources fournissent plusieurs versions du génome humain qui sont les plus utilisées. UCSC propose les versions hg18/hg19 et le GRC fournit GRCh36 et GRCh37. Ces deux génomes de référence sont les plus largement utilisés. UCSC (hg) et GRC (GRCh) ensembles humains sont identiques, mais différent quant à leur nomenclature. Les technologies de séquençage augmentent constamment la longueur des reads nécessitant des algorithmes nouveaux et améliorés. La première génération d'aligneurs a souvent été optimisée pour l'alignement sans trous alors que les programmes d'aujourd'hui peuvent traiter des reads plus longs et les trous de séquences.

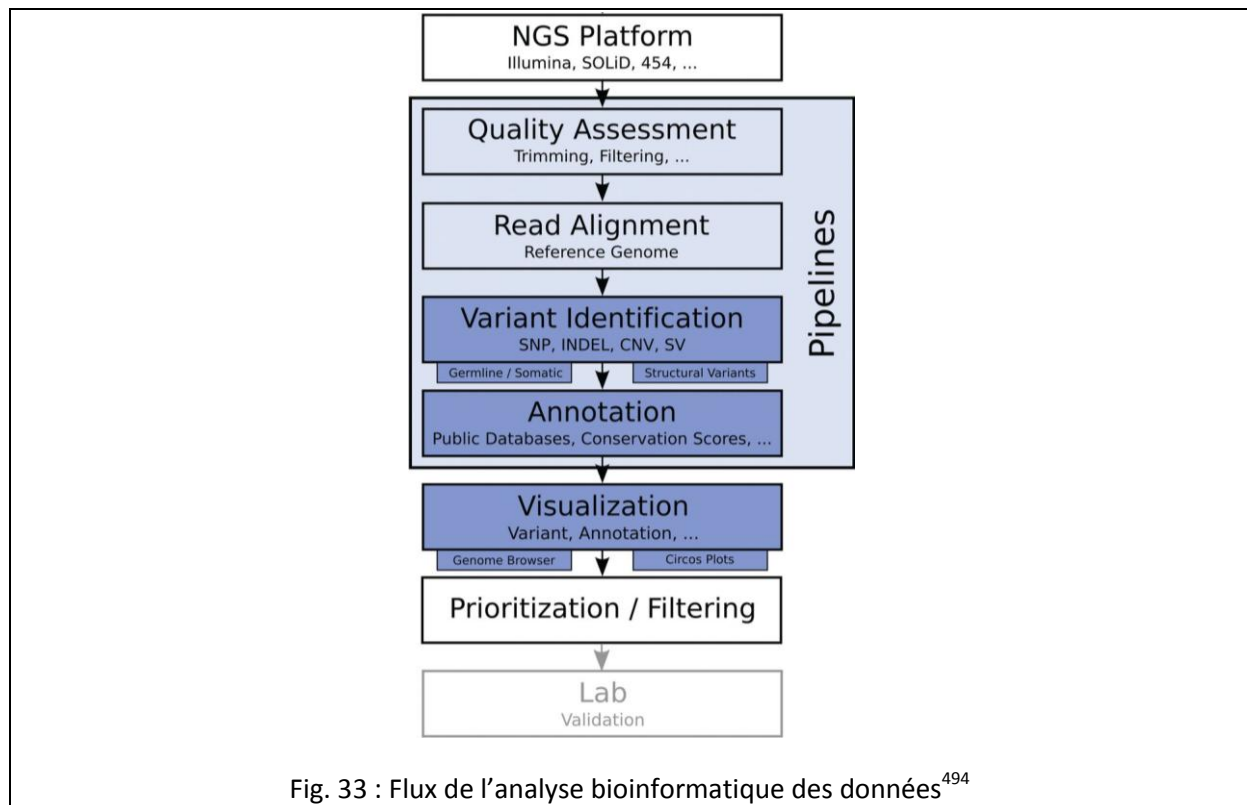


Fig. 33 : Flux de l'analyse bioinformatique des données<sup>494</sup>

Une partie cruciale de l'analyse de données de séquençage est l'identification des variants. Le choix des stratégies appliquées pour l'appel des variants et l'identification des variants structuraux (inversions, translocations, grands INDELs, CNVs, etc) est finalement lié à l'utilisation des données. Ainsi, la conception d'une étude doit intégrer ces paramètres, car un facteur comme la couverture des reads (nombre de reads pour une position génomique unique) joue un rôle important dans l'identification de variants.<sup>495</sup> Le format des fichiers de sortie le plus couramment utilisé est le bam.

Avec la grande quantité de données produites par le NGS, la possibilité de prédire l'impact fonctionnel des variants de façon automatisée est en train de devenir de plus en plus importante. Les annotations permettent de filtrer et hiérarchiser les mutations pathogènes potentielles pour une analyse plus approfondie. Les outils disponibles mettent en œuvre différentes méthodes d'annotation. La plupart se concentrent sur l'annotation de SNPs et d'INDELs, car ils peuvent être facilement identifiés et analysés. L'annotation des variants structuraux est souvent limitée aux CNVs et reste très récente. La forme la plus commune d'annotation est de fournir des liens vers des bases de données publiques de variants comme dbSNP. En termes de prédiction fonctionnelle des variants,

les outils emploient des approches différentes, allant de la conservation interespèces à l'évaluation de l'impact sur la structure des protéines. Le résultat de l'analyse fonctionnelle est une classification en scores reflétant la probabilité d'un effet délétère.

Une étape importante et difficile dans un pipeline d'analyse de données NGS est la validation et la visualisation des résultats générés. La représentation visuelle des données peut être extrêmement utile pour l'interprétation des résultats obtenus. Par conséquent, les outils de visualisation NGS devraient aider les utilisateurs en affichant les reads alignés, la qualité de l'alignement, et des mutations identifiées combinées avec des annotations de diverses sources publiques. Ces outils peuvent se présenter sous la forme de navigateur du génome intégrant les données expérimentales et différents types d'annotations ou de logiciels permettant de visualiser des comparaisons de séquences obtenues de différents individus ou même d'organismes différents.

Le véritable défi de la génétique de nouvelle génération est de réduire la liste des variants candidats et d'interpréter les variants restants dans un contexte biologique.<sup>496</sup> Une approche largement utilisée pour réduire sensiblement la liste des candidats est d'exclure les variants connus qui sont présents dans les bases de données de SNPs, comme on suppose que les variants communs n'ont pas d'effet fort sur le phénotype.<sup>497</sup> Une autre façon de filtrer les données est l'utilisation de familles, à savoir un séquençage d'individus distants avec un phénotype d'intérêt pour identifier la mutation responsable.<sup>498</sup> Cependant, puisque chaque génération introduit jusqu'à 4,5 mutations délétères<sup>499</sup>, lors de séquençage de trio (l'enfant atteint versus les parents non atteints) il est difficile de savoir si une mutation de novo est à l'origine de la maladie. Toutes les méthodes de priorisation ont en commun le risque de suppression du variant pathogène<sup>497</sup> qui se traduit par des taux élevés de faux positifs et de faux négatifs des méthodes de prédiction<sup>500,501</sup> qui ne sont pas suffisamment fiables pour en déduire un diagnostic définitif.<sup>502</sup> L'utilisation de différentes approches de priorisation en combinaison avec des études fonctionnelles sur des cellules ou des animaux semble être la meilleure stratégie pour identifier la cause génétique de la maladie étudiée.

## II. Présentation des travaux

---

### A. Article 1: Whole-Exome Sequencing and High Throughput Genotyping Identified *KCNJ11* as the Thirteenth MODY Gene

Bonnefond A\*, **Philippe J\***, Durand E, Dechaume A, Huyvaert M, Montagne L, Marre M, Balkau B, Fajardy I, Vambergue A, Vatin V, Delplanque J, Le Guilcher D, De Graeve F, Lecoer C, Sand O, Vaxillaire M, Froguel P.

PLoS One. 2012;7(6):e37423. doi: 10.1371/journal.pone.0037423. Epub 2012 Jun 11.

En 2007, quand il était encore trop cher de séquencer des génomes humains entiers, il a été démontré que l'on pouvait capturer la séquence des exons de génomes humains individuels par hybridation sur une puce.<sup>503,504</sup> Ce séquençage a été suivi par la première analyse bioinformatique en profondeur d'un « exome », défini comme l'ensemble des exons d'un génome. L'exome représente ~1% du génome total d'un individu, code l'ensemble des protéines exprimées dans les différents types cellulaires du corps humain et arbore environ 85% des mutations causant des maladies<sup>505</sup>. Dès 2009, le séquençage de 12 exomes humains est publié<sup>506</sup>, suivi par l'utilisation du séquençage de l'exome pour déterminer la base génétique de nombreuses maladies, dont les syndromes de Bartter<sup>505</sup>, de Miller<sup>507</sup> et de Kabuki<sup>508</sup>. Depuis, la chute du coût du séquençage a largement démocratisé cette technique qui reste toutefois onéreuse pour de grandes études.

L'arrivée de ces nouvelles techniques de séquençage au laboratoire a permis de rendre possible des projets basés sur la technique de WES dans le but d'élucider les étiologies génétiques du diabète transmis au sein de familles MODY dont la cause génétique n'est pas connue (MODY-X). Comme le WES peut potentiellement identifier un grand nombre de variants nucléotidiques dans chaque exome séquencé, différentes stratégies d'analyses sont utilisées pour faciliter l'identification des mutations causales : sélection de plusieurs patients non apparentés et affectés par une maladie génétique sévère ou sélection de familles présentant plusieurs sujets atteints d'une maladie donnée. Dans notre cas, 4 membres de la famille ont été sélectionnés : un patient atteint (IV4), ses deux parents - l'un atteint (III5) et l'autre non atteint (III6) - et un cousin atteint (IV5). (**Fig. 34**) L'exome de ces 4 individus a été ensuite capturé par hybridation en phase liquide à l'aide d'un kit Agilent SureSelect 38 Mb à l'exception du IV5 qui a bénéficié d'un kit 50 Mb. Ces 4 exomes sont ensuite séquencés sur le GAIIx d'Illumina. Deux paramètres sont d'intérêt lors d'un séquençage :

- la longueur des reads : dans notre cas 76 bp
- single-end ou paired-end (notre cas) qui permet de séquencer les deux extrémités d'un fragment d'ADN dans un sens et dans l'autre (forward et reverse)

Les données brutes issues du séquençage ont été analysées dans notre pipeline informatique : l'alignement et l'identification des mutations étaient réalisés via Casava, l'annotation via une API fournie par Ensembl et les prédictions fonctionnelles via dbNSFP. Une stratégie d'analyse à plusieurs étapes a été développée pour filtrer les mutations « intéressantes » (c'est-à-dire codantes non-synonymes ou entraînant un gain de codon STOP ou un décalage du cadre de lecture), présentes chez les trois patients atteints et absentes chez le sujet apparenté non atteint, et enfin non

répertoriées dans des bases de variants nucléotidiques (publiques et en collaboration). Une stratégie complémentaire incluant les combinaisons de mutations entre individus a été également mise en place. En effet, l'uniformité de la capture et du séquençage n'est pas homogène entre les individus. Ainsi, un génotype peut être connu chez un individu ou plusieurs et non connu chez les autres, faute d'une profondeur de séquençage insuffisante. Après filtrage et combinaisons, il restait 324 mutations potentiellement causales à évaluer. Ces mutations ont été testées au sein de la famille complète, et chez ~400 sujets contrôles par une technique de génotypage par multiplexage (BeadXpress; Illumina). Ce protocole réalisable en deux jours permet de génotyper sur une même plaque 96 jusqu'à 384 SNPs. Cette technique ne permet de confirmer des variants que dans certaines conditions imposées par le design et le taux d'échec reste élevé. Le séquençage Sanger est une alternative longue et coûteuse, mais souvent utilisée par faute de meilleure option. Environ 8% des mutations n'ont pas pu être testées par BeadXpress et ont dû être testées par séquençage Sanger.

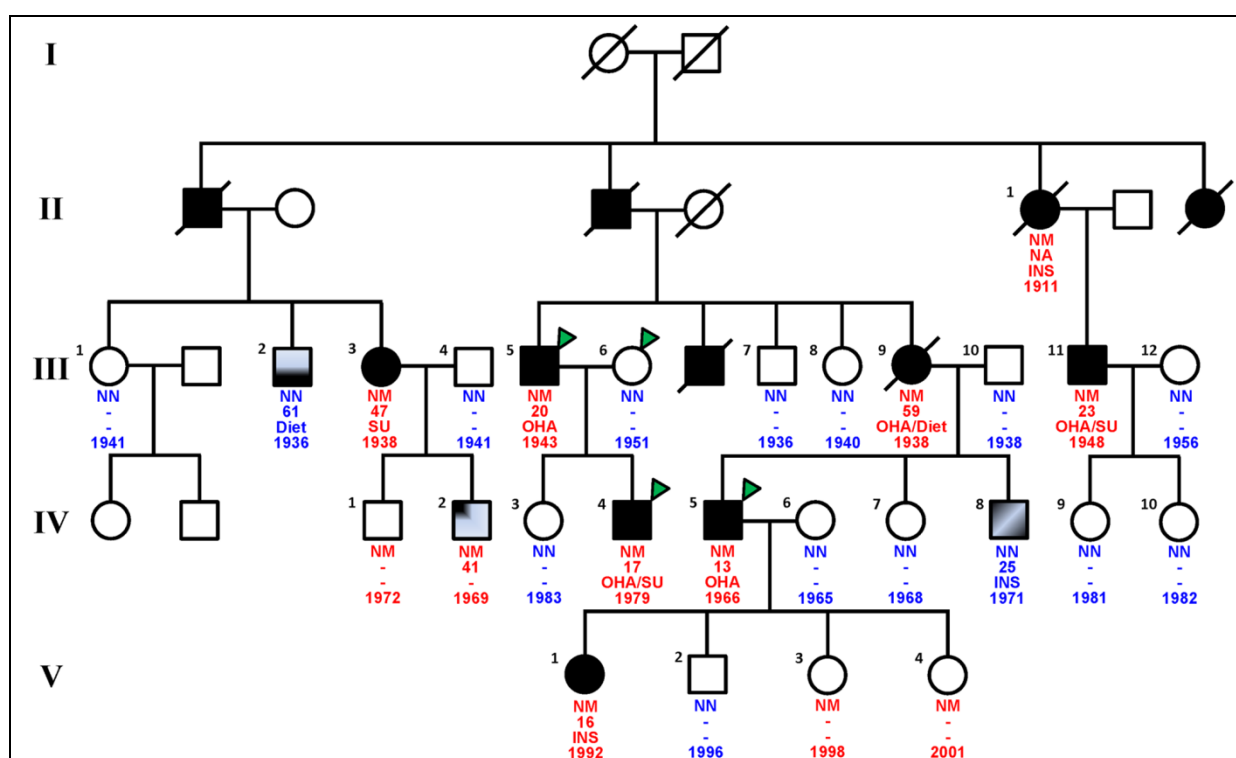
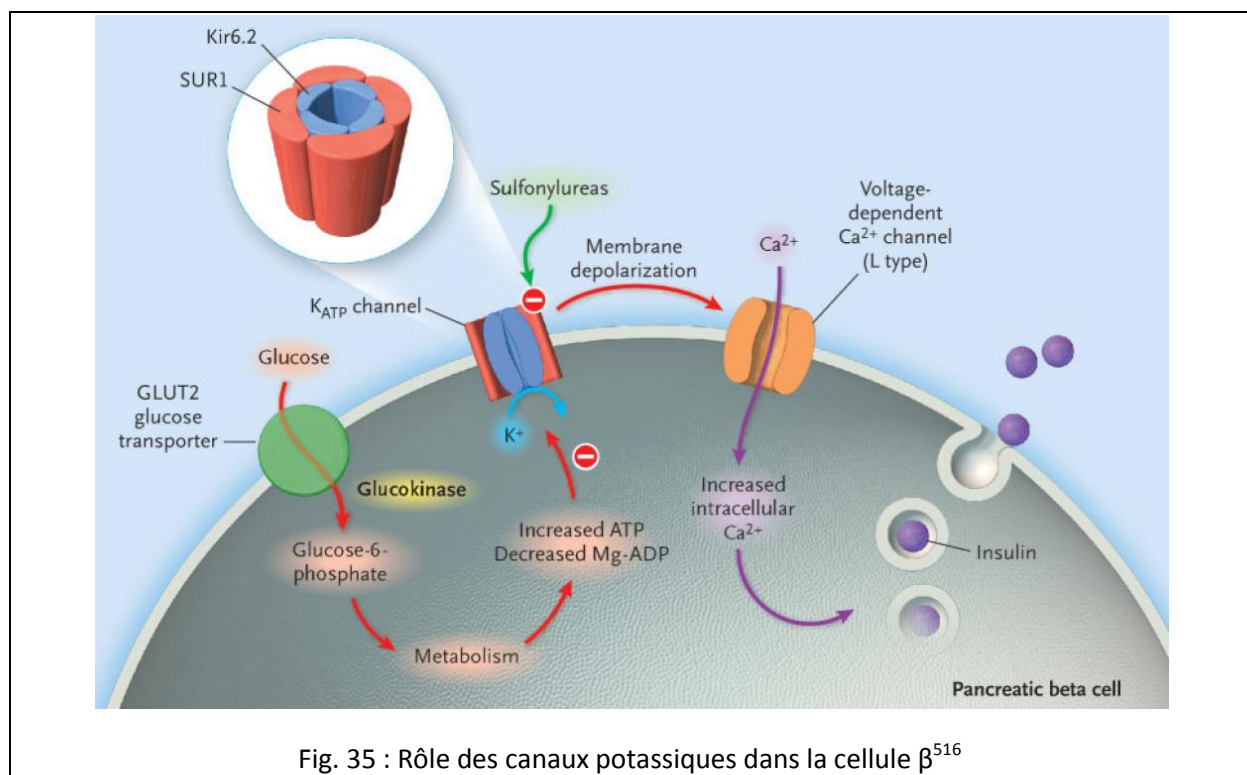


Fig. 34 : Arbre généalogique de la famille MODY étudiée

Sous chaque patient : statut génétique, âge de diagnostic, traitement et année de naissance. NM indique la présence de la mutation p.Glu227Lys dans *KCNJ11* et NN l'absence de la mutation. Les cercles représentent les individus de sexe féminin et les carrés ceux de sexe masculin. Une barre oblique à travers le symbole indique que l'individu est décédé. Les symboles noirs indiquent un DT2 et celui avec une diagonale blanche un DT1. Les symboles remplis à moitié ou au quart indiquent un patient avec IGT et IFG. INS, insuline; OHA, agents hypoglycémiques oraux; SU, sulfonylurée.

Parmi toutes les mutations retrouvées par le WES, une seule mutation était portée par un maximum de diabétiques de type 2 et par un minimum de membres non atteints. C'est une mutation faux sens (p.Glu227Lys) qui concerne le gène *KCNJ11* codant une sous-unité des canaux potassiques. Cette découverte avait un sens par rapport à la littérature, car cette mutation avait déjà été publiée dans des cas de TNDM et avait été montrée comme fonctionnelle dans des ovocytes de *Xenopus*.<sup>509</sup>

Les canaux potassiques sont constitués d'une structure en octamère avec 2 types de sous-unités : 4 sous-unités SUR1 (codé par *ABCC8*) récepteurs aux sulfonyles et 4 sous-unités Kir6.2 (codé par *KCNJ11*) composant le cœur du canal. Quand ce canal est fonctionnel, la production d'ATP par la glycolyse et le cycle de Krebs conduit à l'inhibition et la fermeture des canaux potassiques. Cette fermeture entraîne une dépolarisation de la membrane qui active les canaux calciques et qui permet *in fine* de déclencher l'exocytose de l'insuline. (**Fig. 35**) Dans notre cas, la mutation réduit la sensibilité du canal potassique à l'inhibition par l'ATP et augmente ainsi la probabilité que le canal reste ouvert. Les sulfonyles rétablissent cette sensibilité en se liant aux sous-unités SUR1 du canal et en permettant ainsi leur fermeture. Ce traitement personnalisé chez des porteurs de mutations du canal potassique permet de passer de l'insuline à une simple prise orale quotidienne d'un comprimé et est probablement plus efficace en permettant aux patients d'avoir un meilleur équilibre glycémique.<sup>510</sup> L'individu III3 est un parfait exemple de patient bien équilibré avec presque 30 ans de traitement par sulfamides. Il est intéressant de noter que le traitement aux sulfonyles est aussi utilisé comme traitement du DT2 commun, et que des variabilités interindividuelles de la réponse à ce traitement dépendent de polymorphismes dans les gènes *ABCC8* et *KCNJ11*, mais aussi dans les gènes *TCF7L2*, *IRS1*<sup>511</sup> et *HNF1A*<sup>512</sup>. Le gène *ABCC8* avait déjà été montré comme étant à la fois une cause de NDM<sup>150</sup> et une cause de diabète MODY<sup>513-515</sup>, ce qui n'était pas le cas du gène *KCNJ11*. Ce spectre phénotypique peut être dû à plusieurs facteurs : la sévérité de la mutation, l'épistasie (c'est-à-dire l'interaction avec d'autres gènes), ou encore une influence de l'environnement. À l'inverse des sulfamides, le diazoxide se fixe sur le domaine SUR1, active le canal potassique des cellules  $\beta$ -pancréatiques, entraîne une hyperpolarisation de la membrane, et inhibe la sécrétion d'Insuline. Cette propriété peut être utilisée dans le cas d'un traitement d'un insulinome, une tumeur du pancréas entraînant une production excessive d'insuline.



Cette étude a permis de tirer plusieurs leçons. Le bon phénotypage des patients est un point crucial pour le NGS, à tel point que certains parlent du besoin d'un phénotypage nouvelle



génération.<sup>517</sup> Dans notre cas, l'individu IV8 avait été identifié à tort comme un DT2 au lieu d'un DT1 ce qui avait faussé l'analyse de liaison et donc l'identification de la cause du MODY en 1998. Le diabète de type 2 étant une maladie fréquente, la probabilité d'avoir des phénocopies dans une grande famille MODY n'est pas négligeable. L'individu III2 est par exemple intolérant au glucose, mais ce prédiabète semble plus lié à son âge et son surpoids. De plus, le spectre phénotypique est assez large et la mutation n'est pas totalement pénétrante, car les individus IV1 et IV2 sont porteurs, mais non diabétiques.

En dehors des aspects médicaux, les aspects techniques ne sont pas en reste. À cause des limitations techniques à la fois au niveau de la capture et du séquençage, on observe des disparités entre différentes zones séquencées. Ces problèmes d'uniformité de couverture peuvent être dus à des zones répétées du génome, très pauvres ou très riches en GC, ou des structures secondaires complexes formées par l'ADN. Ces problèmes tendent à diminuer avec le temps au fur et à mesure des nouvelles versions de kits de capture et par l'amélioration de la fiabilité du séquençage (augmentation de la longueur des reads, nouvelle chimie, etc.).

Les challenges concernant l'analyse informatique ne sont pas en reste : selon les algorithmes utilisés pour le trimming<sup>518</sup>, la recherche de SNPs et d'indels<sup>519</sup> ou même leur annotation<sup>520</sup>, le résultat sera différent. Quoi qu'il en soit, les taux de faux positifs et faux négatifs peuvent être minimisés pour les SNPs et les indels mais les CNVs et autres modifications complexes de structure restent mal détectés. Le problème majeur est que les variants obtenus par NGS demandent une confirmation ; or il n'existe pas vraiment de méthode haut débit adaptée.

Ainsi, à chaque étape, l'incertitude de passer à côté de la cause génétique reste entière, d'autant plus que la cause cherchée peut se trouver ailleurs que dans l'exome. Si on estime que 85% des mutations causales sont localisées dans l'exome, les autres peuvent être localisées dans des zones non codantes telles que les UTRs, le promoteur qui contiennent des sites de fixation de facteurs de transcription, d'amplificateurs (enhancer) ou d'autres régulateurs tels que les ARNs non codants.

Le NGS n'en est qu'à ses balbutiements et les développements technologiques et bioinformatiques restent à l'ordre du jour. Ces améliorations passent par un séquençage de meilleure qualité et plus uniforme, de nouveaux algorithmes informatiques incluant des méthodes statistiques<sup>521</sup> ainsi que des bases de données plus complètes pour les SNPs par exemple. L'objectif ultime est d'avoir un minimum de candidats pour des études fonctionnelles sur l'animal ou sur des cellules dans le but de comprendre les mécanismes en jeu et de trouver un traitement.

## B. Article 2: Highly sensitive diagnosis of 43 monogenic forms of diabetes or obesity through one-step PCR-based enrichment in combination with next-generation sequencing

Bonnefond A\*, **Philippe J\***, Durand E, Muller J, Saeed S, Arslan M, Martínez R, De Graeve F, Dhennin V, Rabearivelo I, Polak M, Cavé H, Castaño L, Vaxillaire M, Mandel JL, Sand O, Froguel P.

Diabetes Care. 2014 Feb;37(2):460-7. doi: 10.2337/dc13-0698. Epub 2013 Sep 16.

Le séquençage Sanger est la technique de référence pour un diagnostic moléculaire depuis plus d'une décennie. Sans évolution majeure technologique, les laboratoires utilisent depuis longtemps ce procédé qui pose pourtant des problèmes concrets. Le coût du diagnostic reste important, par exemple plus de 2800\$ pour du NDM (limité au séquençage d'*ABCC8* et *KCNJ11*)<sup>522</sup>, et les temps d'attente sont très longs et peuvent facilement atteindre un an. De plus, pour des raisons d'économie et de temps, le séquençage est effectué de façon séquentielle, gène par gène, ce qui impose d'avoir une hypothèse de départ par rapport au phénotype observé. L'arrivée du NGS a permis d'envisager des alternatives sérieuses au séquençage Sanger. Le WES reste possible pour du diagnostic<sup>523</sup>, mais n'est pas exempt de défaut avec par exemple un séquençage non uniforme, un prix de l'ordre de 1000\$ n'incluant pas l'analyse informatique, et des problèmes éthiques liés aux découvertes fortuites<sup>524</sup>. Dans le but d'évaluer une nouvelle méthode d'enrichissement basée sur une amplification par PCR en microgouttelettes lipidiques, 40 patients dont la cause génétique était déjà connue ont été reséquencés pour preuve de concept.

Diabète			Obésité	
néonatal syndromique	néonatal	MODY	syndromique	non syndromique
	<i>PDX1</i>		<i>BBS1-16</i>	<i>LEP</i>
<i>FOXP3</i>	<i>ABCC8</i>		<i>ALMS1</i>	<i>LEPR</i>
<i>CEL</i>	<i>GCK</i>		<i>GNAS</i>	<i>BDNF</i>
<i>WFS1</i>	<i>INS</i>			<i>MC4R</i>
<i>EIF2AK3</i>	<i>KCNJ11</i>			<i>POMC</i>
<i>NEUROG3</i>	<i>NEUROD1</i>			<i>SIM1</i>
<i>RFX6</i>		<i>HNF1A</i>		<i>NTRK2</i>
<i>PTF1A</i>		<i>HNF1B</i>		<i>PCSK1</i>
<i>GLIS3</i>		<i>HNF4A</i>		

Fig. 36 : Gènes ciblés

La première étape de cette démarche a été la création d'un panel diabète et obésité via la société Raindance. Ce panel inclut 970 couples de primers qui ciblent 43 gènes de susceptibilité aux formes monogéniques de diabète et d'obésité. (**Fig. 36**) Cette technique est unique par l'utilisation de microgouttelettes lipidiques qui permet un enrichissement sans biais. En effet, chaque gouttelette contient un seul couple de primers et un fragment d'ADN génomique d'environ 3kb. Les constituants de la PCR permettent une amplification sans biais allélique et en minimisant les interactions non spécifiques. Chaque ADN de patient est ainsi enrichi séparément pour amplifier les 43 gènes ciblés. Les bibliothèques générées ont ensuite été multiplexées pour être séquencées sur un HiSeq2000 en 76-bp et en paired-end. Les reads obtenus par séquençage avec cette technologie sont dits chimériques et

nécessitent un traitement bioinformatique particulier par l'aligneur BWA-SW<sup>525</sup> (maintenant remplacé par BWA-MEM).

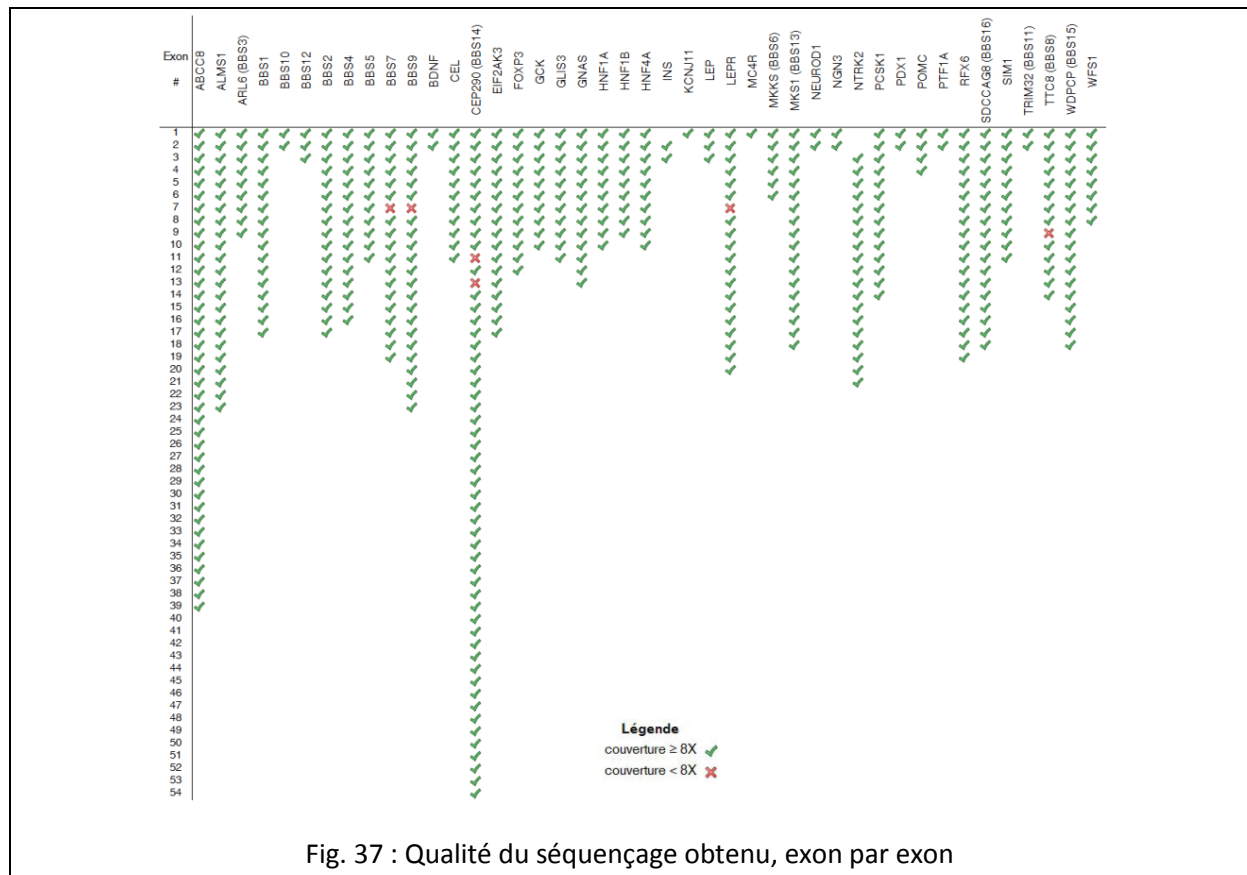


Fig. 37 : Qualité du séquençage obtenu, exon par exon

D'un point de vue qualitatif, les résultats de séquençage étaient très bons avec toutefois 6 exons partiellement séquencés. (**Fig. 37**) Ces exons sont systématiquement mal séquencés chez tous les individus et ne sont pas particulièrement riches ou pauvres en GC mais présentent plutôt une structure particulière avec de nombreuses répétitions de type homopolymères. Ces répétitions sont problématiques pour la plupart des techniques de séquençage existantes y compris celle d'Illumina. En pratique, cela n'a pas nui à la réidentification des mutations causales incluant 23 mutations non synonymes, 7 mutations faux sens, 4 variants touchant les sites d'épissage et 10 indels. Toutes les mutations ont pu être réidentifiées sauf une qui concerne un réarrangement complexe. En visualisant les données, on s'aperçoit que cette mutation est bien présente et résulte en un trou de séquence. (**Fig. 38**) Après investigation, la détection de cette mutation requiert des algorithmes relativement récents de calling dont la dernière version de GATK développée par le Broad Institute. Cette mutation à l'état homozygote peut aussi être détectée par un simple contrôle qualité en vérifiant les zones non séquencées pour chaque individu. Cependant, à l'état hétérozygote, son identification aurait été encore plus problématique et il aurait fallu par exemple utiliser des algorithmes relatifs aux petits CNVs.

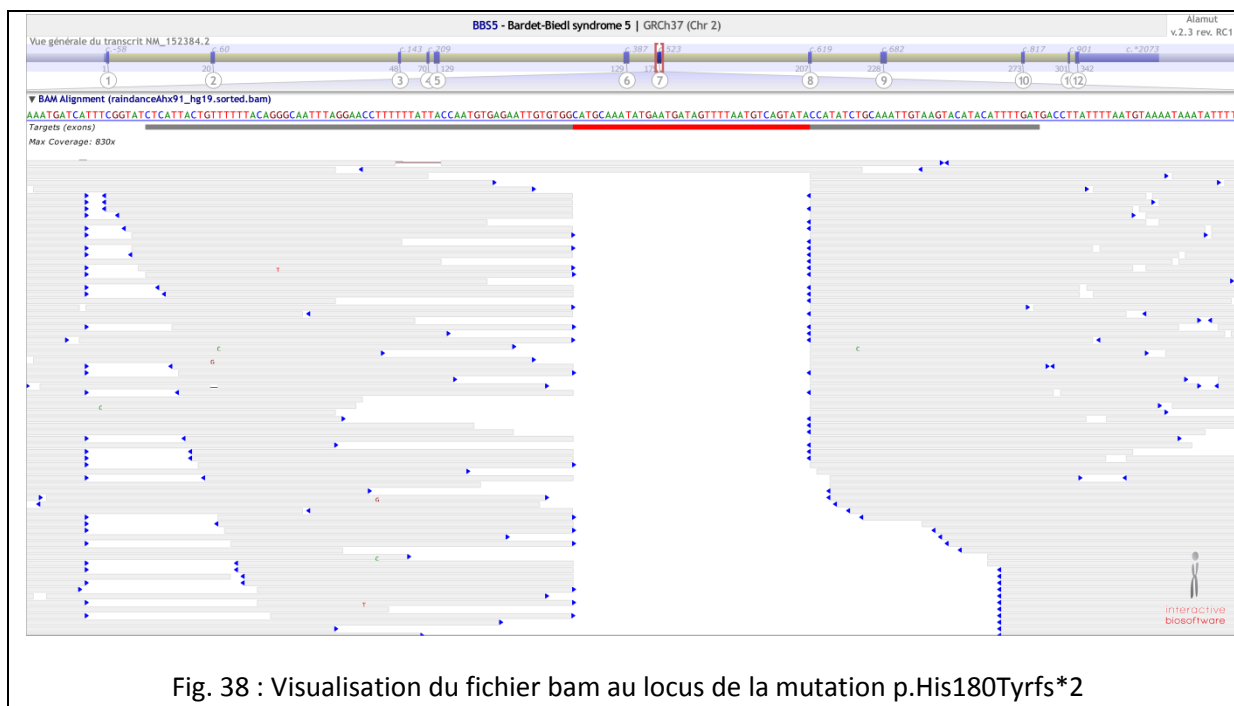


Fig. 38 : Visualisation du fichier bam au locus de la mutation p.His180Tyrfs\*2

Plusieurs enseignements peuvent être tirés de cette étude. Trois patients sont porteurs d'autres mutations potentiellement délétères en plus de celles qu'on attendait. Dans un cas, la deuxième mutation cause le même phénotype que la première. Sans investigation supplémentaire, il est impossible de savoir si les deux mutations jouent un rôle ou seulement l'une des deux. Il s'agit probablement d'un cas rare, mais qui soulève la question de l'approche actuelle par le séquençage Sanger qui est généralement réalisé de façon séquentielle : on s'arrête au premier gène où l'on trouve une mutation. On comprend alors aisément qu'un diagnostic erroné est tout à fait possible, car les cliniciens ne font quasiment jamais d'étude fonctionnelle sur les nouvelles mutations. D'autre part, deux patients obèses ont des mutations dans des gènes du MODY. L'un d'entre eux est un patient atteint de BBS, mais on ne connaît pas le statut glucidique de ce patient. Un diabète ou un prédiabète chez un patient BBS pourrait aisément être mal interprété dans le sens où ce type de syndrome possède un large éventail phénotypique. Chez l'autre patient obèse, on peut aussi supposer qu'un possible diabète serait mal diagnostiqué, car pour un IMC>40, le risque de diabète est multiplié par 10. Ces patients pourraient être traités par insuline en cas de diabète sévère alors qu'un traitement aux sulfonylurées pourrait suffire.

La difficulté d'identification de la mutation causale peut également être due au filtrage. En effet, un certain nombre de mutations sont présentes dans dbSNP, mais sans données de fréquence. Dans notre cas, 14 mutations causales sont présentes dans dbSNP. (**Fig. 39**) Même sans données de fréquence, on peut en déduire leur pathogénicité, car la signification clinique est fournie. Elle provient de ClinVar<sup>526</sup> qui est une base de données phénotypique ouverte intégrée à l'interface d'accès dbSNP. Cependant, dans un cas l'information est erronée, car la mutation est considérée comme bénigne et dans 4 autres cas aucune signification clinique n'est fournie. Ce simple fait impose la prudence lors du filtrage des mutations ou l'intégration de bases phénotypiques externes comme HGMD<sup>527</sup>.

Pour conclure, cette technique d'enrichissement par microgouttelettes lipidiques couplée au NGS est une bonne alternative au séquençage Sanger classique. La qualité des résultats n'est pas

aussi parfaite, mais on peut aisément compléter les amplicons manquants par un séquençage Sanger. La conversion pour le diagnostic clinique risque d'être assez longue, car cette technique d'enrichissement et le NGS demandent une certaine expertise et un coût d'investissement de départ dans les machines et la formation du personnel aussi bien biologique que bioinformatique. Cependant, les économies d'échelle seront importantes avec un coût bien inférieur au séquençage Sanger et au WES grâce aux réactifs moins chers, mais aussi grâce à un volume réduit des données et donc une diminution du coût de stockage. De plus, ce type de méthode évite les problèmes éthiques liés aux découvertes fortuites, un facteur important en clinique. Enfin, la majorité des formes monogéniques de diabète et d'obésité ne sont pas diagnostiquées avec par exemple plus de 80% des cas pour le MODY.<sup>528</sup> Il existe donc un vrai marché et une rentabilité économique<sup>529</sup> qui, on peut l'espérer, se traduiront directement par des améliorations au niveau du patient dont des traitements personnalisés à la clé.

Gène	Mutation	HGMD	dbSNP	
			Présence	Signification clinique
<i>ABCC8</i>	p.Leu225Pro	oui	oui	pathogénique
<i>INS</i>	p.Arg89Cys	oui	oui	pathogénique
<i>ABCC8</i>	p.Leu213Arg	oui	oui	pathogénique
<i>INS</i>	p.Cys96Tyr	oui	oui	pathogénique
<i>HNF1A</i>	p.Gly31Asp	oui	oui	pathogénique
<i>GCK</i>	p.Glu279*	oui	oui	pathogénique
<i>MKKS</i>	p.Tyr37Cys	oui	oui	pathogénique
<i>BBS1</i>	p.Met390Arg	oui	oui	pathogénique
<i>BBS2</i>	p.Asn70Ser	oui	oui	pathogénique
<i>WFS1</i>	p.Ala671Val	oui	oui	vraisemblablement bénin
<i>HNF1B</i>	p.Val61Gly	oui	oui	Non connue
<i>PCSK1</i>	p.Thr175Met	oui	oui	Non connue
<i>MC4R</i>	p.Ala244Glu	oui	oui	Non connue
<i>POMC</i>	p.Glu214Gly	oui	oui	Non connue

Fig. 39 : Mutations causales présentes dans dbSNP

## C. Article 3: A nonsense loss-of-function mutation in *PCSK1* contributes to dominantly inherited human obesity

**Philippe J\***, Stijnen P\*, Meyre D, De Graeve F, Thuillier D, Delplanque J, Gyapay G, Sand O, Creemers JW, Froguel P, Bonnefond A.

Int J Obes (Lond). 2014 Jun 3. doi: 10.1038/ijo.2014.96.

À ce jour, les GWAS ont identifié avec succès plus de 80 loci différents associés à l'obésité polygénique ou des traits relatifs à l'obésité.<sup>129</sup> D'autre part, les formes monogéniques ou du moins les formes d'obésité sévère à début précoce peuvent être dues à des défauts dans 12 gènes. Une partie seulement de ces gènes contribue à la fois à l'obésité monogénique et polygénique : *BDNF*, *DYRK1B*, *MC4R*, *PCSK1*, *POMC* et *SH2B1*. Dès lors, on peut hypothétiser que des mutations rares dans les gènes issus des GWAS jouent un rôle dans l'obésité familiale. Dès l'arrivée du NGS et des premières captures en phase liquide, on pouvait déjà envisager de tester plusieurs centaines de gènes sur plusieurs centaines d'individus.

Gène	Obésité monogénique	Obésité polygénique	Médiane de couverture [1er quartile – 3e quartile]
<i>BDNF</i>	x	x	291 [171 - 411]
<i>CADM2</i>		x	218 [77 - 352]
<i>ETV5</i>		x	205 [124 - 311]
<i>FAIM2</i>		x	87 [45 - 136]
<i>FTO</i>		x	241 [127 - 347]
<i>GNPDA2</i>		x	270 [189 - 347]
<i>GPRC5B</i>		x	228 [99 - 364]
<i>KCTD15</i>		x	130 [64 - 240]
<i>LEP</i>	x		201 [78 - 293]
<i>LEPR</i>	x		343 [222 - 445]
<i>LRP1B</i>		x	281 [192 - 377]
<i>MAF</i>		x	234 [74 - 379]
<i>MAP2K5</i>		x	116 [66 - 241]
<i>MC4R</i>	x	x	341 [260 - 419]
<i>MTCH2</i>		x	243 [143 - 359]
<i>MTIF3</i>		x	209 [56 - 339]
<i>NEGR1</i>		x	339 [221 - 428]
<i>NPC1</i>		x	221 [128 - 334]
<i>NRXN3</i>		x	287 [184 - 398]
<i>NTRK2</i>	x		306 [166 - 421]
<i>NUDT3</i>		x	243 [121 - 357]
<i>PCSK1</i>	x		280 [161 - 393]
<i>POMC</i>	x	x	42 [16 - 101]
<i>PRKD1</i>		x	259 [155 - 374]
<i>PTBP2</i>		x	326 [206 - 422]
<i>QPCTL</i>		x	69 [41 - 115]
<i>SEC16B</i>		x	188 [99 - 313]
<i>SH2B1</i>	x	x	63 [35 - 105]
<i>SIM1</i>	x		318 [141 - 451]
<i>SLC39A8</i>		x	232 [67 - 370]
<i>TFAP2B</i>		x	225 [108 - 352]
<i>TMEM18</i>		x	157 [78 - 278]
<i>TNKS</i>		x	272 [167 - 403]
<i>TNNI3K</i>		x	264 [140 - 402]

Fig. 40 : Liste des gènes d'obésité monogénique et polygénique

C'est ainsi que notre laboratoire en partenariat avec le Génoscope a réalisé une étude pilote sur 13 familles obèses soit 201 individus. Le datamining de plusieurs bases de données a conduit à un classement pondéré et à la sélection de 230 gènes. Ces bases de données incluent les gènes connus de l'obésité monogénique non syndromique, des gènes issus du Mammalian Phenotype Browser (base de données phénotypique de modèles KO et transgéniques de rongeurs), des GWAS (essentiellement GIANT), des gènes inclus dans la délétion 16p11.2<sup>417</sup>, des gènes issus de notre base interne GOOD (gènes exprimés dans les adipocytes et l'hypothalamus), ainsi que des loci issus des études de liaison et répliqués dans l'obésité humaine. D'un point de vue expérimental, les 230 gènes ont été capturés par hybridation en phase liquide à l'aide d'un kit Agilent SureSelect Custom DNA et les librairies séquencées sur un GAllx en 100 bp et en paired-end.

Dans un premier temps, l'analyse des données a été restreinte à 34 gènes connus de l'obésité monogénique et polygénique présents dans notre panel. (**Fig. 40**) La médiane de couverture est supérieure à 100X ce qui est qualitativement très correct sauf pour 4 gènes (*FAIM1*, *POMC*, *QPCTL* et *SH2B1*) qui sont moins bien couverts.

Parmi les variants d'intérêt que nous avons retrouvés (**Fig. 41**), trois d'entre eux n'étaient pas présents chez des individus normopondéraux et seulement chez des obèses : p.Arg80\* dans *PCSK1*, p.Arg81His dans *CADM2* et p.Leu98Pro dans *QPCTL*. Les deux dernières mutations sont pertinentes, car prédites comme délétères ; cependant, les porteurs des familles sont restreints en nombre ce qui ne permet pas de conclure sur une possible coségrégation. La mutation non-sens dans *PCSK1* était portée par une partie des membres obèses de la famille 1 (**Fig. 42**) (#I2, #II1 et #III1) et n'est pas portée par des membres non obèses ce qui va dans le sens d'une coségrégation. Seul l'individu #III3 pose problème, mais en l'absence d'autres données phénotypiques (suivi de la courbe de poids par exemple), on peut supposer qu'il s'agit d'une phénocopie. De façon intéressante, le patient #I2 est aussi atteint de DT2 et le #II1 est prédiabétique, mais on ne sait pas si les porteurs sont touchés par des diarrhées. Comme la mutation non-sens est présente au début du gène *PCSK1*, on peut supposer qu'elle est fonctionnelle et donc que l'enzyme produite ne peut pas remplir son rôle.

Gène	Variant	SNP connu	FAM*	Porteurs hétérozygotes			Nombre de familles touchées	Prédiction de la sévérité**	Ref
				Obèse	Surpoids	Poids normal			
<i>CADM2</i>	c.248G>A / p.Arg81His	NC	NC	2	0	0	1	PD	NC
<i>PCSK1</i>	c.238C>T / p.Arg80*	NC	NC	3	0	0	1	PD	NC
<i>PCSK1</i>	c.539A>G / p.Asn180Ser	NC	NC	0	0	2	1	PD	530
<i>POMC</i>	c.430T>C / p.Phe144Leu	rs201408477	0.001	0	1	2	1	PD	531
<i>QPCTL</i>	c.293T>C / p.Leu98Pro	NC	NC	1	0	0	1	D	NC

Fig. 41 : Liste des variants rares d'intérêt

D, délétère; PD, probablement délétère; PI, probablement inoffensif; NC, Non connu; SNP, single nucleotide polymorphism

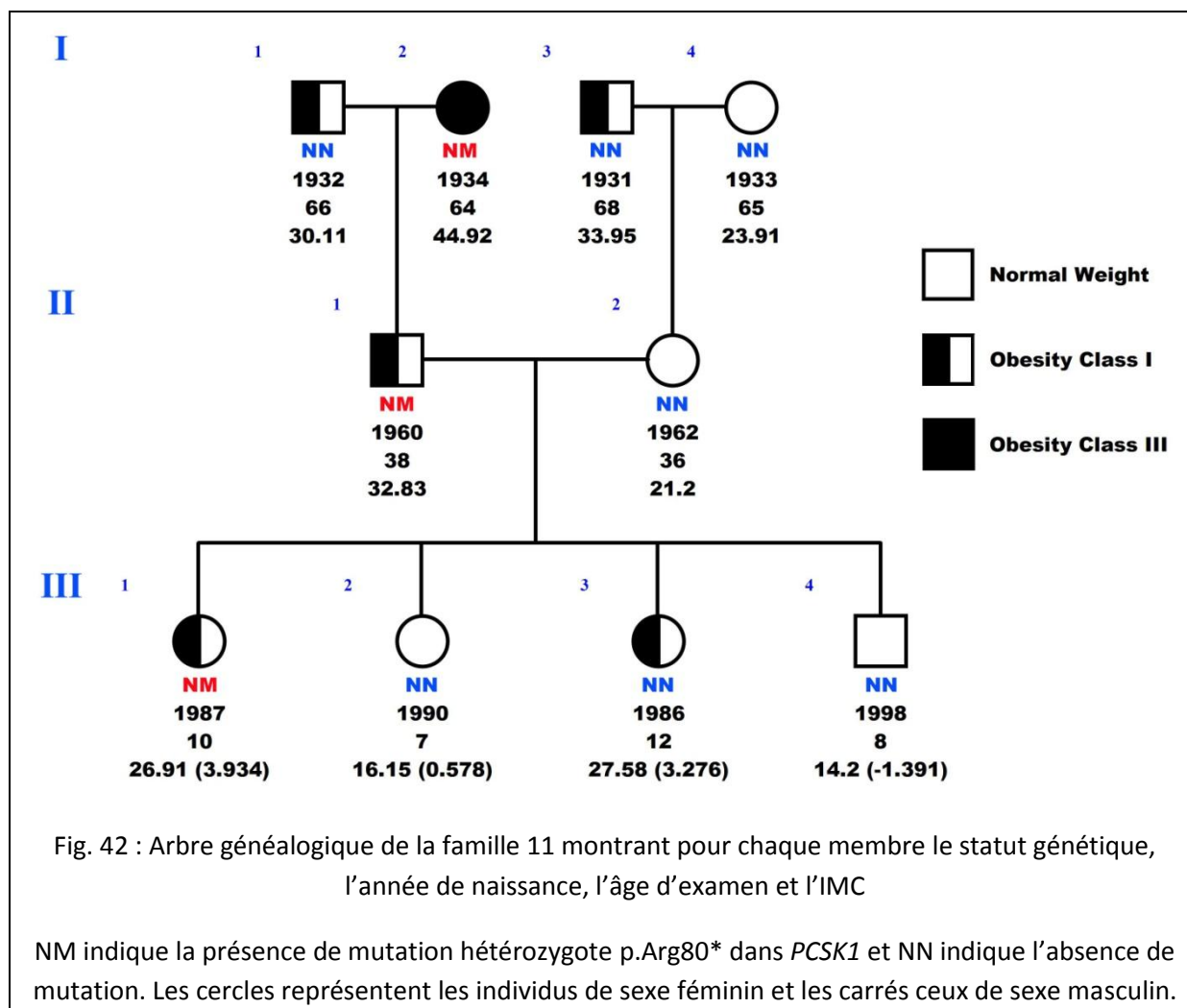
\*FAM selon dbSNP137

\*\*Prédiction in silico de l'effet de chaque variant selon dbNSFP



Ces résultats sont à mettre en parallèle des connaissances sur ce gène. Dans le monde, environ 20 personnes sont homozygotes ou hétérozygotes composites ont une déficience complète en *PCSK1* et ont un certain nombre de symptômes : diarrhée, retard de croissance pendant la petite enfance associée avec un taux de mortalité élevé, obésité sévère à début précoce, polyphagie, diabète insipide, hypogonadisme, hyperproinsulinémie et d'autres dysfonctions endocrines.<sup>426,434,532-</sup>

<sup>534</sup> Aucun cas de déficience hétérozygote conduisant à des symptômes similaires n'est référencé dans la littérature.



D'autre part, le gène *PCSK1* code pour PC1/3 qui fait partie de la famille des proprotéines convertases. Ces protéines subissent un clivage autocatalytique de leur chaperone dans le RE (réticulum endoplasmique). Ce complexe protéique quitte ensuite le RE pour être maturée dans le Golgi ou post-Golgi où un second clivage de la chaperone permet d'activer la proenzyme. (Fig. 43) Dans le cas de PC1/3, le propeptide est clivé à la position 110 dans le RE par un processus intramoléculaire et est dissocié de PC1/3 dans le Golgi grâce à un second clivage C-terminal en position 80 ou 81.<sup>535,536</sup> Le propeptide en *trans* (avec PC1/3 sauvage) inhibe l'enzyme native par une liaison lente et de forte affinité.<sup>537</sup> La question est donc de savoir si le fragment propeptide PC1/3-R80X peut inhiber l'enzyme PC1/3 sauvage. Cette cinétique enzymatique a été évaluée en utilisant un substrat fluorogénique. L'activité de PC1/3 sauvage a diminué de 20% avec PC1/3-R80X et de 25% avec PC1/3(1-110). La maturation et la sécrétion de PC1/3 sauvage en présence ou absence du

propeptide partiel ou complet a été évaluée par western blot dans des lysats de cellules HEK293T. En présence du propeptide partiel ou complet, la quantité de PC1/3 sauvage est augmentée dans le lysat. Cette augmentation pourrait être due au fait que le propeptide maintient l'enzyme sauvage dans le RE ou ralentit le transport intracellulaire par exemple. Tous ces résultats vont dans le sens du phénotype dominant observé dans cette famille et l'inhibition partielle pourrait expliquer qu'une partie seulement des symptômes cliniques sont présents en comparaison des cas de déficience homozygote.

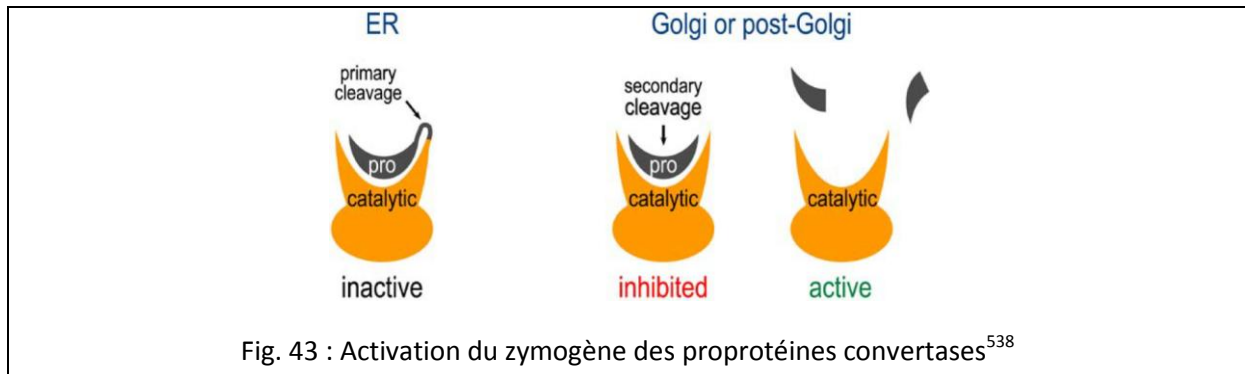


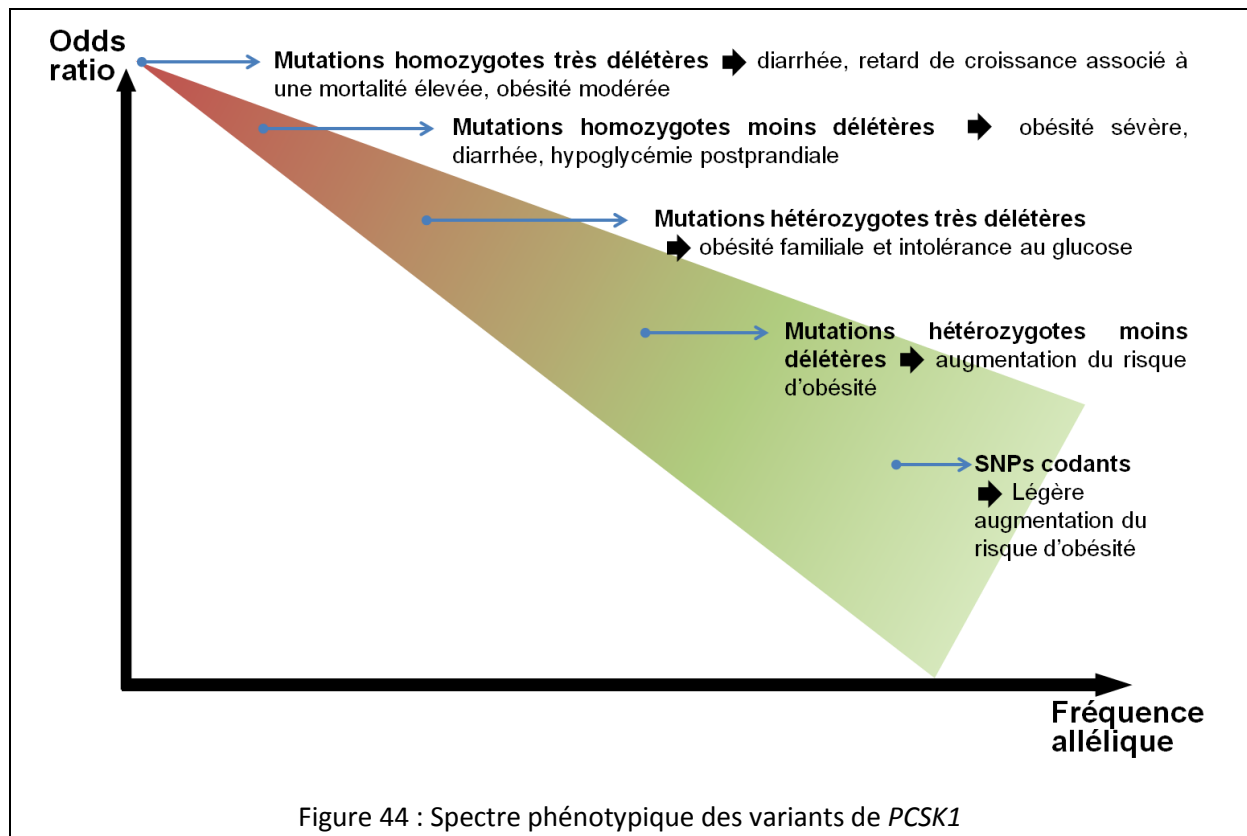
Fig. 43 : Activation du zymogène des proprotéines convertases<sup>538</sup>

On peut ainsi faire le lien entre génotype et phénotype et résumer la situation de la façon suivante (**Fig. 44**):

- Les mutations hétérozygotes composites ou homozygotes entraînent diarrhée, retard de croissance, obésité à début précoce et d'autres caractéristiques cliniques graves
- Une mutation hétérozygote non-sens qui cause une forme dominante d'obésité familiale associée au diabète/prédiabète
- Des mutations hétérozygotes rares augmentant significativement le risque d'obésité<sup>438</sup>
- Des mutations hétérozygotes fréquentes associées à une augmentation modérée du risque d'obésité et de la glycémie et insulinémie à jeun<sup>539,540</sup>

La vraie force des études familiales en comparaison des études cas/contrôles est de pouvoir évaluer la vraie contribution des variants rares par la ségrégation familiale. Par exemple, la mutation p.Asn180Ser dans *PCSK1* augmentait le risque d'obésité dans un cas/contrôle publiée précédemment par notre équipe. Dans notre étude, cette mutation n'a été retrouvée que chez des individus de poids normal, ce qui a tendance à modérer sa contribution à l'obésité sévère.

D'autre part, les mutations rares sont souvent publiées pour une seule famille. Même avec une étude fonctionnelle, on peut prouver le caractère délétère d'une mutation de façon isolée, mais pas dans le contexte biologique de l'individu. Ainsi, la mutation p.Phe144Leu dans *POMC*, publiée précédemment et portée par un père et son fils, altère la liaison entre POMC et MC4R. Cependant, le mode de transmission est normalement autosomique récessif pour *POMC* ce qui remet en doute la contribution de cette mutation par rapport au phénotype observé. Dans notre cas, les 3 individus porteurs ne sont pas obèses donc il est peu probable que cette mutation soit la cause directe de l'obésité dans la première étude.



En conclusion, ce type d'étude familiale permet de confirmer ou d'infirmer la contribution de variants déjà publiés. Ce type d'étude permet aussi de trouver de nouveaux variants dont certains ont un rôle à jouer dans l'obésité comme c'est le cas pour *PCSK1*. Habituellement, un seul probant est séquencé par famille et non la famille complète, ce qui peut être un problème si le probant est mal choisi. Dans notre cas, on peut directement évaluer les résultats sans biais sur la famille complète. Enfin, les 196 gènes restants, dont les résultats ne sont pas encore publiés dans cette étude, laissent présager la découverte de nouveaux loci liés à de nouvelles voies de signalisation impliquées dans la physiopathologie de l'obésité. En effet, aucun variant des 34 premiers gènes ne coségrège dans les 12 familles restantes ce qui implique que la cause génétique de leur obésité est encore à établir. Cette cause peut être monogénique ou polygénique avec un rôle de l'environnement. On peut envisager une approche sans hypothèse par un séquençage du génome éventuellement couplé à un séquençage épigénétique pour élargir le nombre de loci ciblés et analyser les interactions avec l'environnement.

# III. Discussion générale et perspective

---

## A. Le séquençage d'exome

Au cours des dernières années, le NGS a accéléré la détection des variants génétiques conduisant à la découverte rapide de nouveaux gènes associés à diverses maladies. Les success-stories sont nombreuses dont la majorité identifie la cause de la maladie, mais assez peu permettent également un traitement pour le patient. Citons l'exemple d'un américain de 14 ans qui a souffert d'encéphalite aiguë pendant des semaines sans que les médecins parviennent à en trouver la cause.<sup>541,542</sup> Tous les tests de détection de pathogènes étant négatifs et l'état du patient empirant, une méthode expérimentale impliquant le NGS lui a été proposée. Après la détection de la bactérie du genre *Leptospira*, un simple traitement à la pénicilline G a permis le rétablissement complet du garçon. Un autre exemple frappant est le séquençage d'exomes de 12 patients dont la cause génétique n'était pas connue.<sup>543</sup> La moitié des patients a pu être diagnostiqué et pour l'un d'entre eux, l'identification du gène *NGLY1* non connu comme une cause de la maladie a permis le recrutement de 8 autres patients avec un phénotype similaire<sup>544</sup> et de suggérer un traitement potentiel<sup>545</sup>.

Si on se focalise sur les formes monogéniques de diabète et d'obésité, le séquençage d'exome a permis de mettre en évidence les gènes *GATA6*, *PCBD1* et *EEA1* pour le diabète et les gènes *DYRK1B*, *TUB*, *LZTFL1*, *BBIP1* pour l'obésité. Que ce soit pour le diabète et l'obésité ou d'autres maladies, le taux de succès du WES est proche des 25%.<sup>546</sup> Plusieurs raisons sont possibles pour expliquer cet échec :

- Le variant causal est retrouvé, mais est considéré comme non pathogène, car les algorithmes de prédiction protéique le définissent comme tel ou ne sont pas capables de le prédire (création d'un site d'épissage, site régulateur touché, etc.)
- Le variant causal est retrouvé, mais n'est pas annoté correctement faute d'une base de données fiable ; par exemple la fréquence de l'allèle est erronée ou les individus porteurs génotypés avaient le même phénotype, mais on ne le savait pas
- Le variant causal n'est pas reporté lors de l'appel des mutations ; en effet, suivant l'algorithme utilisé, on détecte plus ou moins de SNPs et d'indels ; de plus, faute d'algorithme exploitant les données, certaines variations ne sont pas ou peu analysées (grands indels, inversions, variants structuraux, CNVs, etc.)
- Le variant causal n'est pas couvert avec une profondeur suffisante ; pour des raisons techniques, certaines régions sont mal séquencées, car elles contiennent trop de répétitions, sont riches en GC ou en AT, ou pour d'autres raisons
- Le variant causal n'est pas ciblé ou ne se situe pas dans l'exome ; il y a toujours la possibilité d'un élément régulateur situé dans une région non codante, d'un exon alternatif non connu, d'un gène non connu, d'un variant situé dans le génome mitochondrial, etc.

- Le variant causal n'est pas présent dans le tissu séquencé, par exemple une mutation en mosaïque, c'est-à-dire que tous les tissus ne sont pas touchés
- Une des hypothèses est fausse : le mode de transmission, la pénétrance, un diagnostic erroné, une cause digénique ou même polygénique, une cause épigénétique, un problème de paternité, etc.

Enfin, l'interprétation des résultats dépend des connaissances actuelles. La richesse des données du NGS seul n'est pas suffisante pour comprendre les mécanismes de la pathogenèse des maladies sous-jacentes. Des approches pluridisciplinaires combinant les données de séquençage et les données cliniques avec les connaissances biologiques sont nécessaires pour élucider le rôle des variants dans la génétique humaine. Dans ce contexte, il est crucial que ces données soient liées, organisées et facilement accessibles. Après la déferlante d'« omics », l'heure est à la réconciliation entre ces différentes disciplines. Une analyse combinant les données de génomique, transcriptomique, protéomique, métabolomique ainsi qu'un profil d'autoanticorps a permis chez un individu suivi pendant 14 mois non seulement de prédire le risque relatif à différentes maladies, mais également de suivre en temps réel l'état physiologique du patient.<sup>547</sup> Il a ainsi été possible de prédire la survenue du diabète chez ce patient. Une approche combinée des données de transcriptomique, protéomique et métabolomique a permis d'associer le gène *DHTKD1* au diabète chez la souris puis chez l'homme.<sup>548</sup> Le véritable enjeu est donc l'intégration de différentes données pour au final avoir une meilleure interprétation. Cette vue d'ensemble de plusieurs paramètres est un premier pas vers la médecine personnalisée et également un outil pertinent pour la recherche médicale.

Dans cette optique, de grands projets visent à une compréhension globale du génome. On peut citer l'initiative ENCODE dont les résultats ont notamment été publiés dans 3 articles parus dans *Nature*<sup>549-551</sup> et ont été vivement critiqués par la suite<sup>552</sup>. Le but du projet est de définir les séquences régulatrices du génome, leur impact sur les séquences codantes dans plusieurs tissus humains (180 types cellulaires). Pour ce faire, des techniques telles que l'immunoprécipitation de la chromatine couplée au séquençage et d'autres techniques en lien avec la méthylation de l'ADN ou la fonction de la chromatine. Le pourcentage de séquences non-codantes et fonctionnelles dans le génome humain est estimé à 8,2%.<sup>553</sup> Une autre façon de voir les choses est d'étudier le protéome. Deux équipes ont publié une carte du protéome à l'aide de la spectrométrie de masse.<sup>554,555</sup> Ces cartes ont permis de montrer que pour près de 90% des gènes, on détecte la présence de la protéine associée. Pour les gènes restants, aucune protéine n'est détectée, car les gènes sont probablement inactivés, car devenus inutiles au fil du temps, par exemple des récepteurs liés au goût et à l'odorat. D'autre part, plus de 800 peptides retrouvés n'étaient pas référencés ce qui permet d'améliorer notre compréhension du génome. Le nombre de gènes codant pour des protéines exprimées est en débat constant et pourrait être limité à 19 000.<sup>556</sup> Une autre façon de comprendre le génome est d'étudier le métabolome. Ainsi, un consortium a publié un gwas du métabolome du sang.<sup>557</sup> En utilisant une cohorte de 7800 individus, 145 loci et leurs liens biochimiques ont été associés à plus de 400 métabolites dans le sang humain. Cette approche est intéressante, car le métabolisme est à la fois la résultante d'effets génétiques, mais aussi environnementaux et parce qu'*in fine* cela donne des pistes pour le développement de médicaments et la compréhension de maladies complexes comme le diabète et l'obésité.

## B. Le séquençage ciblé

Le remplacement du séquençage Sanger dans un but diagnostique ne fait aucun doute, la question est plus quelle technologie va le remplacer et dans combien de temps. Le séquençage ciblé offre une alternative attrayante, d'une part à cause du gain de temps et d'argent et d'autre part un apport médical avec une information exhaustive sur tous les gènes. Ce type de séquençage a aussi rendu possible le diagnostic de certains gènes, par exemple le gène de la titine qui compte 364 exons et qui cause des cardiomyopathies isolées, des phénotypes touchant les muscles squelettiques ou des maladies purement infantiles touchant les deux types de muscles striés.<sup>558</sup> La technologie Raindance est une des mieux placées sur ce segment, car les résultats de séquençage offrent une couverture presque parfaite, mais on peut regretter la quantité d'ADN requise qui peut être un facteur limitant. Cependant, les kits salivaires permettent d'obtenir suffisamment d'ADN chez des individus difficiles à prélever, par exemple les prématurés, mais en contrepartie ils ont un impact sur la qualité de l'ADN et donc des résultats.<sup>559</sup>

Quelques hôpitaux en France commencent à utiliser ces techniques de séquençage ciblé, mais l'expertise nécessaire à la fois pour la partie expérimentale et pour la partie bioinformatique limite son développement, d'autant plus que l'investissement n'est pas négligeable et donc que le retour sur investissement se fera sur le long terme. L'expertise nécessaire pour le traitement et le stockage des données est importante, et celle concernant l'interprétation des résultats est encore plus rare. Il est possible pour toutes ces raisons que de la même façon qu'aux États-Unis de grands centres soient créés et traitent les données de toute la France. On peut imaginer des « master panels » qui permettraient de cibler les gènes connus de plusieurs maladies en même temps. L'entreprise GnuBio pourrait prochainement révolutionner le monde du diagnostic NGS dans les petits hôpitaux. Cette entreprise a été récemment rachetée par Bio-Rad. L'un des fondateurs, David Weitz, est à l'origine de la technologie Raindance. Cette technique de séquençage est aussi basée sur l'utilisation de microgouttelettes lipidiques dans laquelle la réaction de séquençage se passe, et permet ainsi d'utiliser environ 1/1000 de la quantité de réactifs dont les autres séquenceurs ont besoin. L'intérêt est simple : le coût est relativement bas et les résultats rapides à obtenir. GnuBio promet un séquenceur à 50 000\$, un coût par échantillon de 200\$ tout inclus, et des résultats obtenus en 3h30 ou moins. Cette technologie qui devrait être commercialisée cette année limite ses panels à 50 gènes probablement pour obtenir rapidement l'accréditation CLIA qui permet d'utiliser la technique dans un environnement clinique. Une autre compagnie qui vise avant tout le diagnostic est la société Roche. Roche a payé 75 millions de dollars Pacific Biosciences pour le développement d'un séquenceur basé sur leur technologie pour un diagnostic clinique, mais les détails de cette collaboration ne font pas légion. Différents acteurs sont intéressés par le marché de la recherche qui est déjà juteux, mais très concurrentiel, et le marché du diagnostic développe encore plus d'intérêt, car les enjeux financiers sont colossaux pour ces multinationales.

Aujourd'hui, la question de la technique à utiliser se pose. En effet, on peut séquencer et analyser un génome en 50 heures.<sup>560</sup> De plus, le coût diminue sans cesse et s'approche des 1000\$ par génome. Un séquençage de génome (WGS) a aussi l'avantage de ne pas faire d'hypothèse sur les gènes d'intérêt, ce qui permet de seulement réanalyser les données dans le cas d'une nouvelle découverte. En effet, le séquençage ciblé requiert de créer un panel de gènes plus ou moins évolutif selon la technologie donc on ne peut pas *a posteriori* mettre à jour les résultats. Cependant, le WGS impose plus de données à stocker et la qualité des données reste inférieure au séquençage ciblé pour

une liste précise de gènes. Le WGS pose aussi le problème éthique des découvertes fortuites. De la même façon, le problème peut se poser lors de découverte d'une tumeur bénigne ou maligne par un scanner des poumons par exemple. La différence est que pour le WGS, la tumeur ne serait pas encore apparue, serait asymptomatique ou n'apparaîtrait peut-être jamais ; en d'autres termes, le caractère prédictif n'est pas de 100%. L'American College of Medical Genetics and Genomics (ACMG) a publié un guide de la conduite à tenir et ils conseillent de restreindre l'analyse à 56 gènes mis en jeu dans 24 maladies pour lesquelles un traitement médical efficace existe. Ces découvertes fortuites concerneraient environ 5% des patients.<sup>561</sup> Un sondage de 1500 personnes a révélé que 78% d'entre eux sont intéressés à recevoir les découvertes fortuites liées à un WGS en donnant une forte priorité aux maladies sérieuses ou traitables. Une grande majorité d'experts en génétique pense aussi que ces découvertes fortuites doivent être communiquées au patient.<sup>562</sup> En résumé, si la différence de coût persiste sur le long terme entre WGS et séquençage ciblé, le WGS ne trouvera pas son public en diagnostic. En revanche, si la différence devient minime, on peut imaginer une migration des laboratoires vers le WGS.

Les techniques de séquençage ciblé gardent un intérêt certain pour la recherche. La technologie Raindance a été utilisée pour cribler des gènes de la surdité<sup>563</sup>, de la maladie de Hirschsprung<sup>564</sup>, du cancer<sup>565</sup>, des maladies mitochondriales<sup>566</sup> et des dystrophies musculaires congénitales<sup>567</sup>. Notre première étude a permis de promouvoir notre méthodologie et une collaboration avec le département de génomique de l'Imperial College a porté ses fruits. En effet, une seconde étude a été publiée sur 39 enfants pakistanais.<sup>568</sup> En utilisant le même panel de gènes, cette cohorte a été criblée et 2 patients sont mutés et homozygotes au niveau du gène *LEPR*. Contrairement à ce qui avait été montré avant, à savoir que les patients mutés pour *LEPR* avaient un phénotype moins sévère que pour *LEP*<sup>439</sup>, ici les patients ont un phénotype similaire. Cette cohorte est très spéciale, car les enfants sont pour la plupart consanguins et de ce fait, elle comporte des homozygotes extrêmement rares. L'avantage de ce criblage des gènes connus pour cette étude comme pour la nôtre est de continuer à rechercher la cause génétique pour les patients non identifiés, par exemple par le séquençage d'exome.

## C. Perspectives thérapeutiques

On peut reprocher à la génétique le manque d'applications directes. Toutefois, le monde médical peut être concilié avec celui de la génétique. Une équipe italienne a tenté de personnaliser le régime de patients obèses en utilisant les variants de 19 gènes en lien avec les préférences alimentaires ou le métabolisme. Par exemple, les individus ayant un profil génétique montrant un métabolisme des lipides moins efficace ont eu un régime hypolipidique. Les patients ayant suivi ce régime personnalisé ont perdu 33% de plus en poids que le groupe contrôle, tout en augmentant leur masse maigre.<sup>569</sup> Cette même équipe a aussi identifié des variants dans le gène *KCNA5* qui influent sur la perception du goût et en particulier du sel. Cette découverte pourrait être intéressante pour prévenir le développement de maladies cardiovasculaires et de l'hypertension ou même pour développer des substituts au sel. Ce résultat est en adéquation avec une autre étude qui montre que les individus qui préfèrent les graisses et le sel ont tendance à moins manger de fruits et de légumes.<sup>570</sup> La perception du goût mais aussi l'odorat ont un rôle non négligeable dans la prise alimentaire. Une étude récente a révélé que les humains étaient capables de détecter par l'odeur



différents taux de graisse dans du lait et ce, indépendamment de la consommation journalière.<sup>571</sup> Une autre étude a montré chez des souris qu'un régime riche en graisses est lié à une réduction des capacités olfactives.<sup>572</sup> Ces études ouvrent de nouvelles pistes et posent la question de l'impact d'un régime riche en sucres et de l'exercice physique sur la perception des odeurs.

Depuis la découverte du tissu adipeux brun, les découvertes n'ont cessé de s'enchaîner et une course à l'activation ou la prolifération de ces tissus est engagée. En effet, ces tissus représentent une piste très sérieuse pour le traitement de l'obésité puisqu'ils sont responsables du maintien de la température corporelle et donc d'une partie de la dépense énergétique. De plus, il semble y avoir un lien avec le métabolisme du glucose et la quantité de tissus adipeux bruns ne semble pas figée dans le temps.<sup>373,573</sup> La dissipation de l'énergie par le tissu adipeux brun est dépendante des mitochondries, en particulier de la protéine UCP-1 qui catalyse la fuite de protons à travers la membrane. Une étude récente a montré l'existence de deux types d'adipocytes bruns : le tissu adipeux brun classique et un nouveau tissu appelé tissu adipeux beige ou clair (beige or brite fat) avec un statut intermédiaire entre le tissu adipeux blanc et brun, et davantage présent chez l'adulte.<sup>574</sup> Un certain nombre de régulateurs transcriptionnels clés sont mis en jeu : PRDM16, PGC-1 $\alpha$ , IRF4, C/EBP $\beta$  et FOXC2. En lien avec PGC-1 $\alpha$ , la protéine membranaire Fndc5 est clivée en un polypeptide appelé irisine qui induit la transformation du tissu adipeux blanc en tissu adipeux brun.<sup>575</sup> L'une des isoformes de PGC-1 $\alpha$ , PGC-1 $\alpha$ 4, stimule la croissance des muscles squelettiques en réponse à un exercice physique. Une étude publiée dans *Cell* au mois de juin identifie l'acteur responsable de ce processus.<sup>576</sup> PGC-1 $\alpha$ 4 stimule la sécrétion d'une hormone, meteorin-like (Metrnl), qui est produite et sécrétée par les muscles après un exercice et le tissu adipeux après une exposition au froid. Metrnl permet la « brunisation » du tissu adipeux blanc, augmente la dépense énergétique, et améliore les paramètres métaboliques dont la tolérance au glucose chez des souris obèses et diabétiques. Cette étude ainsi qu'une autre étude parue en même temps<sup>577</sup> souligne le rôle essentiel du système immunitaire qui sert d'intermédiaire via les macrophages, les éosinophiles et les interleukines 4 et 13. Cette découverte renverse le dogme que seul le cerveau via le système endocrinien contrôle le métabolisme énergétique et ouvre de nouvelles pistes pour un traitement de l'obésité et peut-être même du diabète. En effet, une étude a montré qu'un taux plus important de graisse brune améliore l'homéostasie du glucose et augmente la sensibilité à l'insuline.<sup>578</sup> Cette découverte suggère une capacité de la graisse brune à réguler la glycémie et ouvre des pistes de traitements contre le diabète.

À plus court terme, l'objectif est d'améliorer les traitements existants. Une famille de médicaments qui active PPAR $\gamma$ , les thiazolidinediones, a longtemps servi à traiter le DT2 sur la base de leur activité anti-inflammatoire et de la diminution de la résistance à l'insuline.<sup>579</sup> Cependant, cette classe médicamenteuse a été retirée du marché à cause des effets secondaires qui incluent un gain de poids, une rétention d'eau et une insuffisance cardiaque. Une approche prometteuse vise à disséquer les voies régulatrices qui contrôlent les différents sous-types de gènes ciblés par PPAR $\gamma$  dans les cellules, pour diminuer les effets secondaires. Une étude récente a mis en évidence un groupe de facteurs cellulaires qui se lient à l'ADN et aident PPAR $\gamma$  à réguler un type spécifique de gènes cibles incluant des enzymes responsables de la lipolyse.<sup>580</sup> De la même façon, on peut développer de nouveaux traitements à partir de molécules thérapeutiques déjà connues. Des chercheurs ont eu l'idée de créer une molécule qui combine les propriétés de 2 hormones endocrines : le GLP-1 et le GIP.<sup>581</sup> Cette double incrétine semble être plus efficace que les médicaments présents sur le marché pour le contrôle de glycémie et la réduction du poids. Certains

résultats sont similaires à des doses 10 fois plus élevées d'agonistes des récepteurs au GLP-1. Cette efficacité notoire permettait de réduire des effets secondaires tels que les nausées et les vomissements.

De multiples études seront nécessaires pour évaluer la sécurité de nouveaux traitements qui ont un bénéfice, mais aussi un risque. Le resvératrol, un polyphénol présent dans le vin rouge, présente des effets positifs chez le rongeur<sup>582</sup>, mais ces effets restent à démontrer chez l'humain<sup>583</sup>. Ce type de propagande fait que le resvératrol est disponible en complément alimentaire sans que l'on connaisse les effets et modes d'action de la molécule. De fortes doses entraînent des anomalies du pancréas chez le fœtus de primates, ce qui met en garde contre l'utilisation de resvératrol par la femme enceinte.<sup>584</sup> Le cheval de bataille des antidiabétiques est de diminuer les effets secondaires. Une équipe américaine a démontré qu'une injection de la protéine FGF1 chez des souris diabétiques permettait de garder la glycémie sous contrôle et également de rétablir la sensibilité à l'insuline durant 2 jours.<sup>585</sup> Ce traitement n'a pas d'effets secondaires connus tels qu'un gain de poids, des problèmes hépatiques ou cardiaques, une hypoglycémie, etc. Une nouvelle classe de lipides, les FAHFAs (fatty acid hydroxy fatty acids), permettrait de réduire l'inflammation du tissu adipeux et améliorer l'absorption du glucose par le tissu adipeux par le biais du récepteur GPR120.<sup>586</sup> À l'instar des traitements utilisables *a posteriori*, on peut espérer le développement de traitements prophylactiques pour les prédiabétiques par exemple. L'accumulation d'IAPP dans les cellules  $\beta$  conduit à une toxicité, puis une apoptose qui contribue au DT2. L'autophagie est un processus qui nettoie la cellule des molécules toxiques comme l'IAPP, un processus également mis en jeu dans la maladie d'Alzheimer.<sup>587</sup> La découverte de ce mécanisme pourrait conduire à des traitements pour protéger la cellule  $\beta$ .

La piste la plus en vogue à la fois pour le diabète et l'obésité implique le microbiote intestinal. Ce dernier aurait une valeur prédictive pour le DT2.<sup>65</sup> En effet, les patients ayant une faible diversité bactérienne au niveau du microbiote intestinal semblent avoir une augmentation de l'adiposité et de l'inflammation.<sup>588</sup> Cette diversité peut être en partie retrouvée par un régime, ce qui souligne le caractère réversible et la dynamique du microbiote.<sup>589</sup> Plusieurs espèces bactériennes pourraient être en cause : *Akkermansia muciniphila*, *Firmicutes*, *Bifidobacteria*, *Clostridium leptum* et *Clostridium ramosum*<sup>590</sup>. La vraie question est de savoir si ces changements sont une cause ou une conséquence des maladies métaboliques et comment modifier cette flore. Une équipe a montré qu'une protéine bactérienne, la bile salt hydrolase, modifie les acides biliaires dans l'intestin et que sa surexpression diminue le taux de cholestérol et le gain de poids chez les souris.<sup>591</sup> Un médicament en phase de test nommé NM504 et contenant de l'inuline (mélange de polysaccharides), du bêta-glucane et des antioxydants polyphénoliques a été donné à des adultes prédiabétiques durant 4 semaines. Les résultats de l'HGPO étaient significativement meilleurs que ceux prenant un placebo ainsi qu'une sensibilité à l'insuline améliorée et une diminution de la faim. Une autre étude a exploité la bactérie *E. coli* Nissle 1917 qui a déjà été utilisée comme traitement de la diarrhée.<sup>592</sup> Cette souche a été modifiée pour produire un composé lipidique appelé NAPE qui est normalement synthétisé par l'intestin grêle et qui, une fois converti en NAE, réduit l'apport alimentaire et le gain de poids. Certaines pistes indiquent que la production de NAPE pourrait être diminuée chez les individus consommant un régime riche en graisses. Cette bactérie modifiée introduite dans l'eau des souris soumises à un régime riche en graisses a permis pendant au moins 4 semaines de diminuer la prise de nourriture, la graisse corporelle, la résistance à l'insuline et la stéatose hépatique.

Enfin, les travaux sur les cellules souches se sont intensifiés ces dernières années. Dans un premier temps, la recherche vise avant tout à déterminer les facteurs de différenciation pour ensuite soit partir de cellules souches prélevées sur le patient pour les différencier, soit partir de n'importe quel type cellulaire pour dédifférencier les cellules et les redifférencier dans le type voulu. Cette aptitude à obtenir un type cellulaire voulu comme les cellules  $\beta$  est essentielle pour étudier le diabète pour lequel les prélèvements de pancréas sont extrêmement rares. La thérapie génique a un rôle majeur, car c'est l'outil de choix pour une modification efficace à long terme. Dans ce sens, une étude a été publiée sur des cellules endocrines issues de l'intestin et différenciées en un type de cellules proches des cellules  $\beta$  grâce à l'inhibition du facteur de transcription FOXO1. Les cellules sécrètent l'insuline uniquement en réponse au glucose. Tout récemment, une équipe de l'université d'Harvard a réussi à différencier en grande quantité des cellules souches pluripotentes en cellules dont la fonction semble similaire à celle des cellules  $\beta$  humaines.<sup>593</sup> Si les cellules souches et la thérapie génique représentent des outils indispensables pour la recherche, leur utilisation thérapeutique ne verra pas le jour avant un certain temps, car les obstacles techniques, financiers et éthiques sont encore majeurs.

## D. Perspectives technologiques

L'acteur majeur du séquençage de la décennie passée est clairement Applied Biosystems. Cette décennie, Illumina semble bien parti pour dominer le secteur du séquençage nouvelle génération. En début d'année, Illumina a annoncé une nouvelle machine HiSeq X Ten, ou plutôt un pack de 10 instruments pour 10 millions de dollars. Cet instrument est conçu pour séquençer des génomes humains à l'échelle de population. Le communiqué revendique le génome humain à 1000\$ ce qui semble irréaliste au regard des implications de cette affirmation. En effet, il faudrait que les séquenceurs tournent 24h/24 7j/7 pour séquençer 18 000 génomes humains par an. De plus, ce calcul de coût ne prend pas en compte l'infrastructure nécessaire, l'analyse informatique et l'interprétation des données pourtant essentielles, et repose sur un amortissement sur 4 ans. Le pari est risqué pour les acquéreurs sachant que le GAIIx a eu une durée de moins de 4 ans et les HiSeq moins de 2 ans. Cependant, certains instituts ont été rapidement intéressés puisque le communiqué de presse annonçait 3 acquéreurs : Macrogen, le Broad Institute, et le Garvan Institute of Medical Research. De plus, la compagnie Human Longevity Inc., dirigée notamment par J. Craig Venter, a annoncé l'achat de deux HiSeq X Ten pour séquençer 40000 génomes par an dans le but d'étudier le vieillissement chez l'humain. On ne peut également s'empêcher de penser au projet Genomics England qui a pour but de séquençer 100 000 génomes d'ici 2017 et qui devrait être réalisé par Illumina. Depuis, les annonces n'ont cessé de se multiplier ; Illumina prédit 228 000 génomes humains séquençés en 2014 et ce chiffre devrait atteindre 1,6 million en 2017. Tous les grands centres de séquençage et grandes entreprises font l'acquisition de l'HiSeq X Ten : Novogene, le Wellcome Trust Sanger Institute, le New York Genome Center, WuXi PharmaTech, deCODE genetics. Au-delà de ces acquisitions, on peut se poser la question de l'objectif d'Illumina. Étant en position de quasi-monopole, ils créent en plus de grands centres dont les tarifs ne pourront que difficilement être concurrencés. L'avantage pour Illumina est un support localisé qui revient moins cher que d'avoir 10 fois plus de petits clients qui n'ont pas forcément l'expertise pour résoudre tous les problèmes techniques. Leur récent partenariat avec les 3 laboratoires pharmaceutiques Sanofi,

AstraZeneca et Johnson & Johnson pour le développement de tests oncologiques souligne le but du diagnostic en clinique, un marché très juteux.

Le danger est réel de voir un seul acteur dominer le marché avec une politique de prix élevé. Le principal concurrent est Applied Biosystems, le ténor du séquençage Sanger, qui a longtemps dominé le marché. Leur technologie de séquençage nouvelle génération appelée SOLID n'a jamais percé sur le marché et leur séquenceur basé sur la technologie Ion Torrent peine à se faire une place. Suite à la fusion avec Invitrogen en 2008 pour former Life Technologies, Thermo Fisher Scientific a racheté l'entreprise en 2014, ce qui permettra une possible synergie et de plus gros investissements financiers pour développer la technologie. Roche, qui a initialement acheté la technologie 454 basée sur le pyroséquençage, n'a pas suffisamment amélioré la technologie et a préféré se retirer du marché fin 2013. Ils gardent un partenariat avec la société Pacific Biosciences pour développer des produits relatifs au diagnostic clinique. La course technologique est rude, mais celle des prix l'est davantage. Deux entreprises visent à entrer sur le marché du NGS dès 2015. Après avoir acquis Intelligent Bio-Systems en 2012, Qiagen a repoussé la commercialisation de GeneReader de 2014 à 2015. Ils visent une plateforme tout intégrée grâce à leurs nombreux kits déjà disponibles et par le rachat de sociétés orientées sur l'analyse informatique comme CLC bio et Ingenuity. Le Beijing Institute of Genomics, un institut chinois, vise aussi la commercialisation de son séquenceur BIGIS-4 dont la technique est similaire au pyroséquenceur 454. Le but de ce séquenceur est de casser les prix, d'avoir une ouverture privilégiée sur le marché chinois en pleine expansion, et de viser des applications en lien avec la FDA chinoise.

Pacific Biosciences possède une technologie de séquençage en temps réel d'une seule molécule d'ADN. Ce marché de niche a un intérêt principalement pour l'assemblage de novo de petits génomes, car les reads produits par ce séquenceur peuvent atteindre 10kb. Le débit est beaucoup plus faible que les séquenceurs Illumina et le taux d'erreurs est beaucoup plus élevé. En contrepartie, les reads sont plus longs et les erreurs sont aléatoires ce qui permet de les corriger en ayant une profondeur de séquençage suffisante. En effet, les autres technologies ont des profils bien spécifiques d'erreurs et des biais sur les régions riches et pauvres en GC, ce qui n'est pas le cas ici. D'autre part, cette technique permet également de détecter plusieurs modifications de base dont la méthylation de la cytosine et de l'adénine, très fréquente chez les bactéries. L'utilisation de ce séquenceur pour des applications liées à la génétique humaine reste limitée, mais le PacBio a déjà été utilisé avec succès pour séquencer des répétitions CGG dans le syndrome de l'X fragile<sup>594</sup> et pour séquencer le transcriptome sans biais de tissus<sup>595</sup>, mais aussi de cellules souches<sup>596</sup>. Les problèmes majeurs de ce séquenceur en génétique humaine sont le coût du séquenceur et des réactifs, le débit limité et la nécessité d'avoir une salle blanche, car ce séquenceur est très sensible aux contaminations. Le problème de débit est lié aux ZMWs (Zero-Mode Waveguides) qui sont des petits puits dans lesquels la réaction de séquençage se passe. Seul 1/3 de ces ZMWs est utilisable après le chargement, car une partie d'entre eux ne contient pas de polymérase ou plusieurs. Le chargement peut être augmenté jusqu'à 60% en utilisant des nanoadaptateurs d'origami ADNs qui permettent la création d'une nanostructure qui replie l'ADN de façon optimale<sup>597</sup> ou encore en couplant les ZMWs à des nanopores<sup>598</sup>.

L'avenir du séquençage pourrait venir des séquenceurs basés sur des nanopores. Ces nanopores sont un ensemble de trous formant une membrane constituée par une matière inerte comme le graphène<sup>599</sup> ou par des protéines telles que l'hémolysine alpha et la porine MspA. En

appliquant une faible tension sur la membrane, une molécule d'ADN va passer à travers cette membrane et la mesure du courant au niveau du nanopore permet d'identifier une séquence particulière en allant même jusqu'à détecter des modifications de base de type méthylation. Le principal problème est que la molécule d'ADN passe très vite et donc d'avoir le temps de mesurer une différence de potentiel. L'astuce est de mesurer la signature d'un k-mer (par exemple 6 pb) pour la lier à sa séquence. La société Oxford Nanopore est la première en lice sur ce marché. Le MinION est leur premier séquenceur en bêta test et il est à peu près de la taille d'une clé USB. Les premiers tests révèlent un taux d'erreur assez élevé avec des erreurs systématiques.<sup>600</sup> La société pense dans le futur concevoir un produit avec différents types de nanopore de façon à ce que chacun ait un profil d'erreurs spécifiques qui pourront être corrigées lors de l'alignement. La promesse d'un séquençage avec de très longs reads, peu onéreuse et portable semble proche et viendrait directement concurrencer le PacBio. D'autres entreprises investissent dans des technologies similaires relatives au séquençage par nanopores. Ainsi, Roche a racheté Genia en début d'année, prévoit de commercialiser le premier instrument en fin d'année, et a également investi 15 millions de dollars dans Stratos Genomics, une startup qui développe un séquençage par nanopores. De façon similaire, Illumina a acquis des licences d'utilisation relatives au séquençage par nanopores des universités de Washington et d'Alabama. Cette technologie est basée sur la porine MspA<sup>601</sup> alors qu'Oxford Nanopore et Genia utilisent l'hémolysine alpha<sup>602,603</sup>.

Toutes ces techniques possèdent encore un inconvénient majeur : la préparation de la librairie. Aujourd'hui, aucune technique ne permet de s'en affranchir. On est encore très loin de pouvoir séquencer un génome humain en un seul read, sans erreurs, et à partir d'un simple échantillon de sang ou de salive. Une nouvelle société développe un séquenceur qui pourrait répondre à ces problématiques. Le concept semble simple : une molécule d'ADN est immobilisée dans le séquenceur, les bases sont clivées une à une par pyrophosphorolyse et capturées séparément dans des microgouttelettes pour être lues par fluorescence suite à une réaction chimique. (**Fig.45**) Base4 promet un séquençage haut débit avec un taux d'erreurs très faible sans biais au niveau des homopolymères et des reads de très longue taille.

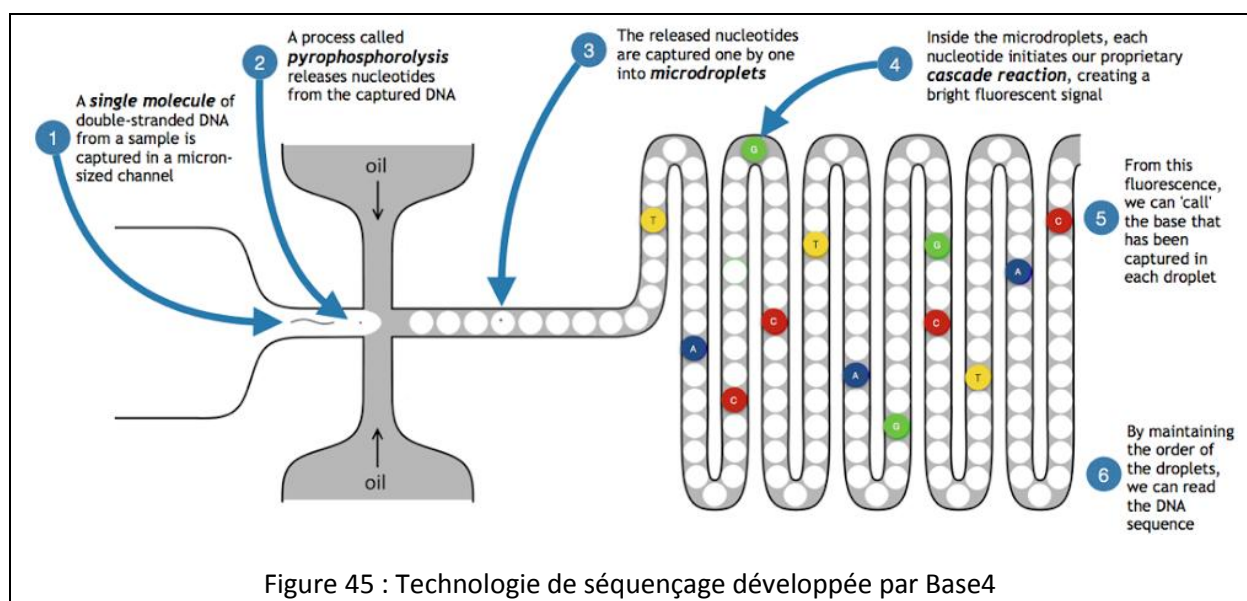


Figure 45 : Technologie de séquençage développée par Base4

## E. Perspectives scientifiques

Le dicton « mieux vaut prévenir que guérir » pourrait être remplacé par « mieux vaut prédire que prévenir », tout du moins pour le diabète et l'obésité. On ne peut en effet que constater l'échec des programmes de prévention devant l'augmentation massive de l'incidence du DT2 et de l'obésité, peut-être faute de moyens, mais aussi d'études sérieuses pour appuyer ces campagnes. À l'inverse, on peut avec ou plus ou moins de justesse prédire le risque de diabète et d'obésité. On peut imaginer un criblage de la population et en particulier des régions à haut risque pour ralentir voire endiguer ces deux épidémies.<sup>604</sup> Les coûts liés au traitement de ces 2 maladies et de leurs complications sont tellement prohibitifs qu'ils valent que l'on s'attarde sur la question. Les outils de diagnostic sont de plus en plus nombreux : un calculateur du MODY<sup>605</sup>, une estimation du risque de devenir obèse pour les nouveaux nés<sup>606</sup>, un biomarqueur du risque de diabète qui est augmenté jusqu'à 12 ans avant le diabète<sup>607</sup>, des biomarqueurs moléculaires comme les miRNAs<sup>608</sup>, etc. Un nouveau biomarqueur, l'hème oxygénase 1 (HO-1), est associé au syndrome métabolique et plus précisément une inflammation chronique et au risque de développer un DT2 chez des patients obèses.<sup>609</sup> Au-delà de l'intérêt prédictif, cette découverte pourrait donner le jour à un possible traitement, en bloquant HO-1. Outre l'aspect purement récréatif, l'algorithme Cortana de Microsoft a prédit sans se tromper l'issue de tous les matchs du mondial de football. En effet, les prédictions dépendent aussi des progrès en intelligence artificielle. Par exemple, le superordinateur Watson d'IBM a réussi à diagnostiquer correctement un cancer des poumons dans 90% des cas contre 50% de réussite pour les médecins.<sup>610</sup> Ces aides informatiques au diagnostic feront probablement partie de l'avenir médical.

L'une des pistes les plus prometteuses pour la prédiction est le génome. Si les gwas ont apporté des avancées majeures dans la compréhension des mécanismes, les SNPs associés au diabète ou à l'obésité n'augmentent que modestement les scores de risque.<sup>611,612</sup> Indirectement se pose la question de l'héritabilité manquante : comment expliquer que la génétique ait un impact majeur, dans les études de jumeaux par exemple, mais que les SNPs fréquents n'expliquent qu'une petite partie de l'héritabilité? Plusieurs explications sont possibles. Le calcul de l'héritabilité sur les jumeaux se base notamment sur les lois de Mendel qui, on le sait maintenant, sont en partie biaisées. De plus, le calcul ne peut pas être exact à cause des interactions gènes-environnement ce qui explique en partie les différentes estimations selon les études. D'autre part, l'héritabilité n'est qu'une mesure parmi d'autres et le degré d'importance d'un variant génétique dépend de la mesure utilisée.<sup>613</sup> Une autre explication se situe dans l'hypothèse même derrière les gwas : des maladies telles que le diabète et l'obésité sont fréquentes, donc des variants fréquents doivent en être la cause. Cette hypothèse rendait l'approche possible via les puces à ADN, mais elle excluait tous les variants avec une MAF<5%. Dorénavant, avec l'arrivée des nouvelles techniques de séquençage, on peut prendre en compte ces variants plus rares. Les variants à découvrir dépendent grandement de la population étudiée et on peut voir par exemple un enrichissement d'une mutation fonctionnelle dans *HNF1A* pour une population latino<sup>614</sup> avec un DT2 ou dans *TBC1D4* chez des groenlandais<sup>139</sup>. D'autres éléments n'ont pas été analysés dans les puces à cause de la technique en elle-même. Par exemple, il existe un certain nombre de zones du génome non couvertes par les tagSNPs et on ne peut pas analyser les petits CNVs et des éléments répétés comme les microsatellites<sup>615</sup>. Encore aujourd'hui, le génome humain n'est pas séquençé à 100% et les CNVs requièrent de nouvelles approches pour être étudiés, par exemple par la technologie NanoString ou par PCR digitale. Un gros



travail d'interprétation des SNPs associés dans les gwas reste aussi à faire via des études fonctionnelles qui exploitent des techniques telles que le ChIP-seq, le FAIRE-seq, le RNA-seq pour identifier les *cis*-eQTLs, et des techniques de capture de conformation des chromosomes (3C, 4C, 5C et Hi-C).

On peut donc s'inquiéter d'initiatives liées à la génomique personnelle (personal genomics) comme 23andMe qui s'adressent à des particuliers et qui prédisent un risque pour telle ou telle maladie. On l'aura compris, les limitations techniques sont suffisamment nombreuses pour bannir ou au moins réglementer ce genre de service. En effet, certaines de ces compagnies vont jusqu'à faire un séquençage d'exome et surtout à l'interpréter. Une expertise est nécessaire pour être conscient des limitations qui sont pour la plupart en phase d'être résolues. Les variants en mosaïque chez les parents des patients atteints peuvent par exemple être détectés et analysés en WES.<sup>616</sup> Ce type de phénomène biologique sera davantage analysé dans un futur proche grâce à l'essor du séquençage « single cell ». Le concept est que l'hétérogénéité cellulaire nécessite qu'on séquence différentes cellules séparément pour comprendre la complexité des échanges. Les premières études se sont focalisées sur le séquençage d'ARN de cellules du système immunitaire ou de cellules cancéreuses.<sup>617,618</sup> Cette technique présente aussi une grande utilité pour comprendre le développement des cellules souches dont le génome semble beaucoup plus instable qu'on ne le pensait jusqu'à présent.<sup>619</sup> D'autres domaines pourraient bénéficier de cette avancée : l'épigénétique, l'étude du microbiome intestinal, le diagnostic prénatal, la police scientifique, etc.

Si les technologies relatives au séquençage s'améliorent de jour en jour, les outils bioinformatiques atteignent enfin une certaine stabilité et maturité. En effet, les débuts du NGS se sont accompagnés du développement d'un nombre très important de programmes bioinformatiques, si bien qu'il était difficile d'avoir une visibilité dans ce domaine. Au fur et à mesure, certains outils sont devenus très populaires et pour la plupart sont développés par de grands instituts : BWA et SAMtools développés par le Wellcome Trust Sanger Institute, GATK et PLINK développés par le Broad Institute, etc. Le nombre d'outils étant déjà important, de plus en plus d'études se penchent sur des comparaisons entre les outils et sur les paramètres à optimiser. Une étude récente a utilisé une forme rare de lignée cellulaire, CHM1, pour évaluer différents outils.<sup>620</sup> CHM1 est une môle hydatiforme qui est le résultat d'une grossesse molaire où l'ovocyte n'a pas de noyau et où le génome du sperme est dupliqué. Cette lignée est idéale, car tous les variants hétérozygotes sont forcément des erreurs. Les séquences brutes montrent un taux d'erreurs de 10% ce qui est très élevé et étonne même l'auteur de l'article. Ces erreurs sont partiellement corrigées par un filtre des régions de faible complexité ou si on croise les résultats de deux pipelines bioinformatiques différentes. Des solutions plus complexes sont évoquées telles que l'assemblage de novo ou utiliser un ensemble de génomes. De façon intéressante, le génome de référence a lui aussi un impact : les références hs37d5 et GRCh38 donnent un meilleur résultat que GRCh37. La référence hs37d5, utilisée par le projet 1000 génomes, contient 35.4Mb de séquences qui sont issues de plusieurs assemblages de novo, mais qui ne sont pas présentes dans GRCh37. La lignée CHM1 a d'ailleurs servi à résoudre certaines régions présentes dans GRCh38, car certaines régions du génome sont tellement variables qu'il est impossible de reconstituer une séquence de référence à partir des deux allèles différents.<sup>621</sup> Les prochaines versions du génome humain utiliseront des résultats de séquençage du génome humain en 54X sur un PacBio. Certains gènes hautement répétitifs, par exemple *MUC5AC*, mais aussi les télomères et centromères peuvent être résolus.<sup>622</sup> L'objectif futur est d'arriver à un génome le plus complet possible appelé « platinum », et pour ce faire, de



séquencer plusieurs échantillons (d'autres ethnies) et d'utiliser le système optique de la société BioNano qui permet d'obtenir une structure de l'ADN pour améliorer encore la fiabilité de la référence. On peut imaginer que d'autres données comme des modifications de bases (méthylation, hydroxyméthylation, etc.) viendront à être intégrées par la suite quand des techniques suffisamment fiables permettront d'obtenir cette information sur le génome entier.

Le projet GENCODE estime à 57820 le nombre de gènes et à 20345 le nombre de gènes codant pour une protéine.<sup>623</sup> Cette complexité est encore plus importante au niveau du transcriptome où on compte 196520 transcrits dont 81814 sont codants. Pire encore, le nombre de gènes codants serait variable d'un individu à l'autre, car dépendant des polymorphismes.<sup>624</sup> Près de 40% des gènes n'ont pas de fonction pour la protéine associée.<sup>625</sup> Certains de ces gènes sont dits orphelins, car ils n'ont aucun homologue parmi les autres espèces animales.<sup>626</sup> Ce problème de taille implique potentiellement des biais importants lors de l'interprétation des résultats de WES par exemple. L'analyse systématique de ces gènes « inconnus » passera par l'étude de modèles animaux (rat, souris, poisson-zèbre, *Caenorhabditis elegans*, etc.) et sera complétée par la bioinformatique, la protéomique et la biologie des systèmes. La découverte du système CRISPR-cas9 chez la bactérie jouera un rôle majeur, car cet outil peut être utilisé pour éditer simplement le génome.<sup>627</sup> On peut sérieusement envisager aujourd'hui de réaliser des banques de cellules inactivées pour chacun des gènes et ainsi explorer leurs fonctions, ou encore d'introduire des mutations en *cis* et en *trans* pour étudier leur impact fonctionnel.<sup>628-631</sup> Les grands projets de séquençage de génomes humains contribuent aussi à déterminer la fonction des gènes. En effet, les variants perte de fonction sont une ressource précieuse qui permet à l'image des modèles animaux d'avoir un équivalent KO humain.<sup>632</sup> Une étude récente a découvert, par le biais de 150 000 individus séquencés, que la perte du gène *SLC30A8* protégeait du DT2 alors que le phénotype inverse est observé chez la souris.<sup>633</sup> Ce type de découverte permet de valider *in vivo* une cible thérapeutique en évitant les biais des études animales. C'est peut-être de cette façon qu'on trouvera de vraies réponses thérapeutiques, en étudiant les variants qui protègent du diabète et de l'obésité. Dans ce même sens, on peut imaginer plus d'études qui étudient des phénotypes miroirs, par exemple la maigreur et l'hypoglycémie, pour trouver des réponses dans les phénotypes opposés. Citons les gènes *ABCC8*, *GCK*, *HNF1A*, *HNF4A*, *KCNJ11* qui peuvent, en fonction des mutations, être une cause de diabète ou d'hypoglycémie. On peut également citer le locus 16p11.2 qui s'il est délété cause une obésité et s'il est dupliqué cause une maigreur.<sup>417,634</sup>

Une des questions fondamentales est de comprendre pourquoi les taux de DT2 et d'obésité sont aussi élevés. En 1962 a été proposée l'hypothèse du génotype économe (thrifty gene hypothesis) d'abord appliquée au diabète puis étendue à l'obésité.<sup>635</sup> Cette hypothèse s'inspire de l'idée de sélection naturelle de Darwin et la base de cette hypothèse est que des mutations auparavant bénéfiques pour la survie sont devenues préjudiciables dans l'environnement actuel. Par exemple, les périodes cycliques de famine et d'abondance de nourriture auraient favorisé les gènes allant dans le sens du stockage de l'énergie. Inversement, l'hypothèse du génotype à la dérive (drifty gene hypothesis) prétend que l'épidémie résulte d'une suppression des pressions de sélection liées aux prédateurs.<sup>636</sup> D'une certaine façon, une dérive génétique aurait eu lieu favorisant la sélection de gènes qui contrôlent la limite haute du stockage de graisses. En effet, des gènes favorisant la vitesse, l'agilité, l'endurance, la capacité athlétique, et la maigreur ne deviennent plus nécessaires pour échapper aux prédateurs. Une nouvelle hypothèse prend en compte l'impact de la mondialisation, l'évolution démographique de la population et les différences ethniques.<sup>637</sup> La pandémie actuelle

serait la résultante de facteurs environnementaux qui ont commencé il y a 70 000 ans quand l'homme a quitté l'Afrique et a migré à travers le monde. Sur la base des différences ethniques, les auteurs proposent que l'adaptation aux différents climats soit la véritable cause. Les gènes relatifs à la thermorégulation seraient en cause, un facteur important lors d'un climat très chaud ou très froid. Cela expliquerait pourquoi les taux d'obésité sont les plus bas chez les Caucasiens, les Japonais, les Chinois et les Coréens et parmi les plus élevés chez les populations africaines. C'est donc peut-être en étudiant la génétique des populations d'un point de vue de l'évolution qu'on trouvera des réponses.<sup>638</sup> Le climat n'est sûrement pas la cause unique. L'exemple du variant dans *TBC1D4* qui augmente significativement le risque de DT2 chez les Groenlandais peut être la résultante d'un changement de régime.<sup>139</sup> En effet, ils se nourrissaient à l'origine d'une majorité de produits issus de la mer riches en protéines et en graisses, mais pauvres en sucres. On peut imaginer que la combinaison d'une alimentation riche en sucre et du variant à risque est la cause de l'augmentation du DT2. Ainsi, une combinaison de facteurs génétiques et environnementaux peut expliquer l'épidémie de diabète et d'obésité et la question est plus de savoir dans quelle mesure chaque facteur contribue à ces maladies. Quoi qu'il en soit, la complexité des mécanismes génétiques et la diversité des facteurs environnementaux seront une source abondante de recherche pour encore de nombreuses années à venir.

## IV. Bibliographie

---

- 1 Loriaux, D. L. Diabetes and The Ebers Papyrus: 1552 B.C. *The Endocrinologist* **16**, 55-56  
10.1097/1001.ten.0000202534.0000283446.0000202569 (2006).
- 2 Bhattacharya, S. Sushrutha - our proud heritage. *Indian journal of plastic surgery : official  
publication of the Association of Plastic Surgeons of India* **42**, 223-225 (2009).
- 3 Willis, T. *Pharmaceutice rationalis, sive Diatriba de medicamentorum operationibus in  
humano corpore, authore Tho. Willis.* (e Theatro Sheldoniano, 1674).
- 4 Bernard, C. *Leçons de physiologie expérimentale appliquée à la médecine: Cours du semestre  
d'hiver 1854-55 : avec 22 figures.* (Baillière, 1855).
- 5 Mering, J. & Minkowski, O. Diabetes mellitus nach Pankreasexstirpation. *Archiv f. experiment.  
Pathol. u. Pharmacol* **26**, 371-387, doi:10.1007/BF01831214 (1890).
- 6 Banting, F. G., Best, C. H., Collip, J. B., Campbell, W. R. & Fletcher, A. A. Pancreatic extracts in  
the treatment of diabetes mellitus: preliminary report. 1922. *CMAJ : Canadian Medical  
Association journal = journal de l'Association medicale canadienne* **145**, 1281-1286 (1991).
- 7 Sanger, F. The free amino groups of insulin. *The Biochemical journal* **39**, 507-515 (1945).
- 8 Adams, M. J. et al. Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **224**, 491-495  
(1969).
- 9 Global Status Report on Noncommunicable Diseases 2010. (World Health Organization,  
2011).
- 10 IDF. IDF Diabetes Atlas, 6th edn. (International Diabetes Federation, Brussels, Belgium:  
International Diabetes Federation, 2013).
- 11 Druet, C. et al. [Type 2 diabetes in France: epidemiology, trends of medical care, social and  
economic burden]. *Presse medicale* **42**, 830-838, doi:10.1016/j.lpm.2013.02.312 (2013).
- 12 Roche. Inserm-TNS Healthcare-Rochep ObÉpi Roche 2009 : Enquête épidémiologique  
nationale sur le surpoids et l'obésité. (Roche, 2009).
- 13 Étude nationale nutrition santé, ENNS. (INVS, 2006).
- 14 Ricci P, C. M., Detournay B, Poutignat N, Kusnik-Joinville O, Raimond V et al, pour le comité  
scientifique d'Entred. Coûts des soins remboursés par l'Assurance maladie aux personnes  
traitées pour diabète : Études Entred 2001 et 2007. *Bull Epidemiol Hebdom* **42-43**, 464-469  
(2009).
- 15 American Diabetes, A. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **37  
Suppl 1**, S81-90, doi:10.2337/dc14-S081 (2014).
- 16 Nagamine, K. et al. Positional cloning of the APECED gene. *Nat Genet* **17**, 393-398,  
doi:10.1038/ng1297-393 (1997).
- 17 Finnish-German, A. C. An autoimmune disease, APECED, caused by mutations in a novel gene  
featuring two PHD-type zinc-finger domains. *Nat Genet* **17**, 399-403, doi:10.1038/ng1297-  
399 (1997).
- 18 Bennett, C. L. et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked  
syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet* **27**, 20-21, doi:10.1038/83713  
(2001).
- 19 Wildin, R. S. et al. X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy  
syndrome is the human equivalent of mouse scurfy. *Nat Genet* **27**, 18-20, doi:10.1038/83707  
(2001).
- 20 Biason-Lauber, A. et al. Identification of a SIRT1 mutation in a family with type 1 diabetes.  
*Cell Metab* **17**, 448-455, doi:10.1016/j.cmet.2013.02.001 (2013).
- 21 Flanagan, S. E. et al. Activating germline mutations in STAT3 cause early-onset multi-organ  
autoimmune disease. *Nat Genet* **46**, 812-814, doi:10.1038/ng.3040 (2014).

- 22 Kadowaki, T. *et al.* Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance. *Science* **240**, 787-790 (1988).
- 23 Yoshimasa, Y. *et al.* Insulin-resistant diabetes due to a point mutation that prevents insulin proreceptor processing. *Science* **240**, 784-787 (1988).
- 24 International Association of, D. *et al.* International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* **33**, 676-682, doi:10.2337/dc09-1848 (2010).
- 25 American Diabetes, A. Standards of medical care in diabetes--2011. *Diabetes Care* **34 Suppl 1**, S11-61, doi:10.2337/dc11-S011 (2011).
- 26 Vandorsten, J. P. *et al.* NIH consensus development conference: diagnosing gestational diabetes mellitus. *NIH consensus and state-of-the-science statements* **29**, 1-31 (2013).
- 27 Kim, C., Newton, K. M. & Knopp, R. H. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care* **25**, 1862-1868 (2002).
- 28 American Diabetes, A. Standards of medical care in diabetes--2014. *Diabetes Care* **37 Suppl 1**, S14-80, doi:10.2337/dc14-S014 (2014).
- 29 Landon, M. B. *et al.* A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med* **361**, 1339-1348, doi:10.1056/NEJMoa0902430 (2009).
- 30 Crowther, C. A. *et al.* Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med* **352**, 2477-2486, doi:10.1056/NEJMoa042973 (2005).
- 31 ADA. *Statistics About Diabetes*, <<http://www.diabetes.org/diabetes-basics/statistics/>> (2011).
- 32 Liew, G., Michaelides, M. & Bunce, C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. *BMJ open* **4**, e004015, doi:10.1136/bmjopen-2013-004015 (2014).
- 33 Morrish, N. J., Wang, S. L., Stevens, L. K., Fuller, J. H. & Keen, H. Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* **44 Suppl 2**, S14-21 (2001).
- 34 Colosia, A. D., Palencia, R. & Khan, S. Prevalence of hypertension and obesity in patients with type 2 diabetes mellitus in observational studies: a systematic literature review. *Diabetes, metabolic syndrome and obesity : targets and therapy* **6**, 327-338, doi:10.2147/DMSO.S51325 (2013).
- 35 Zhou, X. H. *et al.* Diabetes, prediabetes and cancer mortality. *Diabetologia* **53**, 1867-1876, doi:10.1007/s00125-010-1796-7 (2010).
- 36 Emerging Risk Factors, C. *et al.* Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med* **364**, 829-841, doi:10.1056/NEJMoa1008862 (2011).
- 37 Shikata, K., Ninomiya, T. & Kiyohara, Y. Diabetes mellitus and cancer risk: review of the epidemiological evidence. *Cancer science* **104**, 9-14, doi:10.1111/cas.12043 (2013).
- 38 Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **20**, 1183-1197 (1997).
- 39 Syed, I. A. Glycated haemoglobin; past, present, and future are we ready for the change. *JPMMA. The Journal of the Pakistan Medical Association* **61**, 383-388 (2011).
- 40 Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemc and nonglycemc pathways. *Diabetes* **59**, 3229-3239, doi:10.2337/db10-0502 (2010).
- 41 Vaysse, N. Physiologie du pancréas exocrine. *EMC - Hépatogastroentérologie* **2**, 59-74, doi:10.1016/j.emchg.2005.01.002 (2005).
- 42 Bosco, D. *et al.* Unique arrangement of alpha- and beta-cells in human islets of Langerhans. *Diabetes* **59**, 1202-1210, doi:10.2337/db09-1177 (2010).
- 43 Krstic, R. V. *Human Microscopic Anatomy: An Atlas for Students of Medicine and Biology*. (Springer, 1991).
- 44 Wikipedia. *Islets of Langerhans*, <[http://en.wikipedia.org/wiki/Islets\\_of\\_Langerhans](http://en.wikipedia.org/wiki/Islets_of_Langerhans)> (

- 45 Reaven, G. M. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes* **37**, 1595-1607 (1988).
- 46 Kahn, S. E. *et al.* Quantification of the relationship between insulin sensitivity and beta-cell function in human subjects. Evidence for a hyperbolic function. *Diabetes* **42**, 1663-1672 (1993).
- 47 Yi, P., Park, J. S. & Melton, D. A. Betatrophin: a hormone that controls pancreatic beta cell proliferation. *Cell* **153**, 747-758, doi:10.1016/j.cell.2013.04.008 (2013).
- 48 Fujikawa, T. *et al.* Leptin engages a hypothalamic neurocircuitry to permit survival in the absence of insulin. *Cell Metab* **18**, 431-444, doi:10.1016/j.cmet.2013.08.004 (2013).
- 49 Stolarczyk, E. *et al.* Improved insulin sensitivity despite increased visceral adiposity in mice deficient for the immune cell transcription factor T-bet. *Cell Metab* **17**, 520-533, doi:10.1016/j.cmet.2013.02.019 (2013).
- 50 Kahn, S. E., Cooper, M. E. & Del Prato, S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet*, doi:10.1016/S0140-6736(13)62154-6 (2013).
- 51 Rahier, J., Guiot, Y., Goebbels, R. M., Sempoux, C. & Henquin, J. C. Pancreatic beta-cell mass in European subjects with type 2 diabetes. *Diabetes, obesity & metabolism* **10 Suppl 4**, 32-42, doi:10.1111/j.1463-1326.2008.00969.x (2008).
- 52 Poitout, V. & Robertson, R. P. Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocr Rev* **29**, 351-366, doi:10.1210/er.2007-0023 (2008).
- 53 Jurgens, C. A. *et al.* beta-cell loss and beta-cell apoptosis in human type 2 diabetes are related to islet amyloid deposition. *The American journal of pathology* **178**, 2632-2640, doi:10.1016/j.ajpath.2011.02.036 (2011).
- 54 Perl, S. *et al.* Significant human beta-cell turnover is limited to the first three decades of life as determined by in vivo thymidine analog incorporation and radiocarbon dating. *J Clin Endocrinol Metab* **95**, E234-239, doi:10.1210/jc.2010-0932 (2010).
- 55 Dunning, B. E. & Gerich, J. E. The role of alpha-cell dysregulation in fasting and postprandial hyperglycemia in type 2 diabetes and therapeutic implications. *Endocr Rev* **28**, 253-283, doi:10.1210/er.2006-0026 (2007).
- 56 Gromada, J., Franklin, I. & Wollheim, C. B. Alpha-cells of the endocrine pancreas: 35 years of research but the enigma remains. *Endocr Rev* **28**, 84-116, doi:10.1210/er.2006-0007 (2007).
- 57 Drucker, D. J. The biology of incretin hormones. *Cell Metab* **3**, 153-165, doi:10.1016/j.cmet.2006.01.004 (2006).
- 58 Nauck, M. A., Vardarli, I., Deacon, C. F., Holst, J. J. & Meier, J. J. Secretion of glucagon-like peptide-1 (GLP-1) in type 2 diabetes: what is up, what is down? *Diabetologia* **54**, 10-18, doi:10.1007/s00125-010-1896-4 (2011).
- 59 Ward, W. K., Beard, J. C., Halter, J. B., Pfeifer, M. A. & Porte, D., Jr. Pathophysiology of insulin secretion in non-insulin-dependent diabetes mellitus. *Diabetes Care* **7**, 491-502 (1984).
- 60 Stefater, M. A., Wilson-Perez, H. E., Chambers, A. P., Sandoval, D. A. & Seeley, R. J. All bariatric surgeries are not created equal: insights from mechanistic comparisons. *Endocr Rev* **33**, 595-622, doi:10.1210/er.2011-1044 (2012).
- 61 Saeidi, N. *et al.* Reprogramming of intestinal glucose metabolism and glycemic control in rats after gastric bypass. *Science* **341**, 406-410, doi:10.1126/science.1235103 (2013).
- 62 Diamant, M., Blaak, E. E. & de Vos, W. M. Do nutrient-gut-microbiota interactions play a role in human obesity, insulin resistance and type 2 diabetes? *Obesity reviews : an official journal of the International Association for the Study of Obesity* **12**, 272-281, doi:10.1111/j.1467-789X.2010.00797.x (2011).
- 63 Kootte, R. S. *et al.* The therapeutic potential of manipulating gut microbiota in obesity and type 2 diabetes mellitus. *Diabetes, obesity & metabolism* **14**, 112-120, doi:10.1111/j.1463-1326.2011.01483.x (2012).
- 64 Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60, doi:10.1038/nature11450 (2012).

- 65 Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99-103, doi:10.1038/nature12198 (2013).
- 66 Vrieze, A. *et al.* Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143**, 913-916 e917, doi:10.1053/j.gastro.2012.06.031 (2012).
- 67 Jumpertz, R. *et al.* Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am J Clin Nutr* **94**, 58-65, doi:10.3945/ajcn.110.010132 (2011).
- 68 Suez, J. *et al.* Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature*, doi:10.1038/nature13793 (2014).
- 69 Nonogaki, K. New insights into sympathetic regulation of glucose and fat metabolism. *Diabetologia* **43**, 533-549, doi:10.1007/s001250051341 (2000).
- 70 Obici, S., Zhang, B. B., Karkanias, G. & Rossetti, L. Hypothalamic insulin signaling is required for inhibition of glucose production. *Nat Med* **8**, 1376-1382, doi:10.1038/nm798 (2002).
- 71 Lam, T. K., Gutierrez-Juarez, R., Pocai, A. & Rossetti, L. Regulation of blood glucose by hypothalamic pyruvate metabolism. *Science* **309**, 943-947, doi:10.1126/science.1112085 (2005).
- 72 Bass, J. & Takahashi, J. S. Circadian integration of metabolism and energetics. *Science* **330**, 1349-1354, doi:10.1126/science.1195027 (2010).
- 73 Marcheva, B. *et al.* Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466**, 627-631, doi:10.1038/nature09253 (2010).
- 74 Peschke, E. & Muhlbauer, E. New evidence for a role of melatonin in glucose regulation. *Best Pract Res Clin Endocrinol Metab* **24**, 829-841, doi:10.1016/j.beem.2010.09.001 (2010).
- 75 Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study, G. *et al.* Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med* **359**, 1464-1476, doi:10.1056/NEJMoa0805017 (2008).
- 76 Tamborlane, W. V., Sherwin, R. S., Genel, M. & Felig, P. Reduction to normal of plasma glucose in juvenile diabetes by subcutaneous administration of insulin with a portable infusion pump. *N Engl J Med* **300**, 573-578, doi:10.1056/NEJM197903153001101 (1979).
- 77 Knowler, W. C. *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* **346**, 393-403, doi:10.1056/NEJMoa012512 (2002).
- 78 van den Berghe, G. *et al.* Intensive insulin therapy in critically ill patients. *N Engl J Med* **345**, 1359-1367, doi:10.1056/NEJMoa011300 (2001).
- 79 Holden, S. E., Gale, E. A., Jenkins-Jones, S. & Currie, C. J. How many people inject insulin? UK estimates from 1991 to 2010. *Diabetes, obesity & metabolism* **16**, 553-559, doi:10.1111/dom.12260 (2014).
- 80 Inzucchi, S. E. *et al.* Management of hyperglycaemia in type 2 diabetes: a patient-centered approach. Position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia* **55**, 1577-1596, doi:10.1007/s00125-012-2534-0 (2012).
- 81 Barnett, A. H. *et al.* Angiotensin-receptor blockade versus converting-enzyme inhibition in type 2 diabetes and nephropathy. *N Engl J Med* **351**, 1952-1961, doi:10.1056/NEJMoa042274 (2004).
- 82 Klahr, S. *et al.* The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. Modification of Diet in Renal Disease Study Group. *N Engl J Med* **330**, 877-884, doi:10.1056/NEJM199403313301301 (1994).
- 83 Frank, R. N. Diabetic retinopathy. *N Engl J Med* **350**, 48-58, doi:10.1056/NEJMra021678 (2004).
- 84 Shapiro, A. M. *et al.* International trial of the Edmonton protocol for islet transplantation. *N Engl J Med* **355**, 1318-1330, doi:10.1056/NEJMoa061267 (2006).
- 85 Schauer, P. R. *et al.* Bariatric surgery versus intensive medical therapy in obese patients with diabetes. *N Engl J Med* **366**, 1567-1576, doi:10.1056/NEJMoa1200225 (2012).



- 86 Mingrone, G. *et al.* Bariatric surgery versus conventional medical therapy for type 2 diabetes. *N Engl J Med* **366**, 1577-1585, doi:10.1056/NEJMoa1200111 (2012).
- 87 Shai, I. *et al.* Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study. *Diabetes Care* **29**, 1585-1590, doi:10.2337/dc06-0057 (2006).
- 88 Krauss, R. M. Lipids and lipoproteins in patients with type 2 diabetes. *Diabetes Care* **27**, 1496-1504 (2004).
- 89 Lunetta, M. *et al.* Hypertension and related risk factors in type 2 diabetes mellitus. *Panminerva medica* **40**, 169-173 (1998).
- 90 Besingi, W. & Johansson, A. Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol Genet* **23**, 2290-2297, doi:10.1093/hmg/ddt621 (2014).
- 91 Montonen, J. *et al.* Dietary patterns and the incidence of type 2 diabetes. *American journal of epidemiology* **161**, 219-227, doi:10.1093/aje/kwi039 (2005).
- 92 Nelson, K. M., Reiber, G., Boyko, E. J. & Nhanes, III. Diet and exercise among adults with type 2 diabetes: findings from the third national health and nutrition examination survey (NHANES III). *Diabetes Care* **25**, 1722-1728 (2002).
- 93 George, E. S., Rosenkranz, R. R. & Kolt, G. S. Chronic disease and sitting time in middle-aged Australian males: findings from the 45 and Up Study. *The international journal of behavioral nutrition and physical activity* **10**, 20, doi:10.1186/1479-5868-10-20 (2013).
- 94 Fletcher, B., Gulanick, M. & Lamendola, C. Risk factors for type 2 diabetes mellitus. *The Journal of cardiovascular nursing* **16**, 17-23 (2002).
- 95 Spranger, J. *et al.* Inflammatory cytokines and the risk to develop type 2 diabetes: results of the prospective population-based European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. *Diabetes* **52**, 812-817 (2003).
- 96 Dabelea, D. *et al.* Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: a study of discordant sibships. *Diabetes* **49**, 2208-2211 (2000).
- 97 Meisinger, C. *et al.* Sex differences in risk factors for incident type 2 diabetes mellitus: the MONICA Augsburg cohort study. *Arch Intern Med* **162**, 82-89 (2002).
- 98 Kivimäki, M. *et al.* Long working hours, socioeconomic status, and the risk of incident type 2 diabetes: a meta-analysis of published and unpublished data from 222 120 individuals. *The Lancet Diabetes & Endocrinology*, doi:[http://dx.doi.org/10.1016/S2213-8587\(14\)70178-0](http://dx.doi.org/10.1016/S2213-8587(14)70178-0).
- 99 Daousi, C. *et al.* Prevalence of obesity in type 2 diabetes in secondary care: association with cardiovascular risk factors. *Postgraduate medical journal* **82**, 280-284, doi:10.1136/pmj.2005.039032 (2006).
- 100 Schienkiewitz, A., Schulze, M. B., Hoffmann, K., Kroke, A. & Boeing, H. Body mass index history and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. *Am J Clin Nutr* **84**, 427-433 (2006).
- 101 Tushuizen, M. E. *et al.* Pancreatic fat content and beta-cell function in men with and without type 2 diabetes. *Diabetes Care* **30**, 2916-2921, doi:10.2337/dc07-0326 (2007).
- 102 Lee, J. S. *et al.* Clinical implications of fatty pancreas: correlations between fatty pancreas and metabolic syndrome. *World journal of gastroenterology : WJG* **15**, 1869-1875 (2009).
- 103 Pollin, T. I. Epigenetics and diabetes risk: not just for imprinting anymore? *Diabetes* **60**, 1859-1860, doi:10.2337/db11-0515 (2011).
- 104 Hales, C. N. *et al.* Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ* **303**, 1019-1022 (1991).
- 105 Hales, C. N. & Barker, D. J. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* **35**, 595-601 (1992).
- 106 Silverman, B. L., Rizzo, T. A., Cho, N. H. & Metzger, B. E. Long-term effects of the intrauterine environment. The Northwestern University Diabetes in Pregnancy Center. *Diabetes Care* **21 Suppl 2**, B142-149 (1998).
- 107 Oken, E., Rifas-Shiman, S. L., Field, A. E., Frazier, A. L. & Gillman, M. W. Maternal gestational weight gain and offspring weight in adolescence. *Obstetrics and gynecology* **112**, 999-1006, doi:10.1097/AOG.0b013e31818a5d50 (2008).



- 108 Bellamy, L., Casas, J. P., Hingorani, A. D. & Williams, D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet* **373**, 1773-1779, doi:10.1016/S0140-6736(09)60731-5 (2009).
- 109 Pettitt, D. J. *et al.* Association between maternal diabetes in utero and age at offspring's diagnosis of type 2 diabetes. *Diabetes Care* **31**, 2126-2130, doi:10.2337/dc08-0769 (2008).
- 110 Nilsson, E. *et al.* Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes* **63**, 2962-2976, doi:10.2337/db13-1459 (2014).
- 111 El Hajj, N. *et al.* Metabolic programming of MEST DNA methylation by intrauterine exposure to gestational diabetes mellitus. *Diabetes* **62**, 1320-1328, doi:10.2337/db12-0289 (2013).
- 112 Ruchat, S. M. *et al.* Gestational diabetes mellitus epigenetically affects genes predominantly involved in metabolic diseases. *Epigenetics : official journal of the DNA Methylation Society* **8**, 935-943, doi:10.4161/epi.25578 (2013).
- 113 West, N. A., Kechris, K. & Dabelea, D. Exposure to Maternal Diabetes in Utero and DNA Methylation Patterns in the Offspring. *Immunometabolism* **1**, 1-9, doi:10.2478/immun-2013-0001 (2013).
- 114 Wu, C. S., Nohr, E. A., Bech, B. H., Vestergaard, M. & Olsen, J. Long-term health outcomes in children born to mothers with diabetes: a population-based cohort study. *PLoS One* **7**, e36727, doi:10.1371/journal.pone.0036727 (2012).
- 115 Gagnon, C. *et al.* Serum 25-hydroxyvitamin D, calcium intake, and risk of type 2 diabetes after 5 years: results from a national, population-based prospective study (the Australian Diabetes, Obesity and Lifestyle study). *Diabetes Care* **34**, 1133-1138, doi:10.2337/dc10-2167 (2011).
- 116 Pan, A., Lucas, M., Sun, Q. & *et al.* Bidirectional association between depression and type 2 diabetes mellitus in women. *Arch. Intern. Med.* **170**, 1884-1891, doi:10.1001/archinternmed.2010.356 (2010).
- 117 Barnard, K., Peveler, R. C. & Holt, R. I. Antidepressant medication as a risk factor for type 2 diabetes and impaired glucose regulation: systematic review. *Diabetes Care* **36**, 3337-3345, doi:10.2337/dc13-0560 (2013).
- 118 Huth, C. *et al.* Job Strain as a Risk Factor for the Onset of Type 2 Diabetes Mellitus: Findings From the MONICA/KORA Augsburg Cohort Study. *Psychosomatic medicine*, doi:10.1097/PSY.0000000000000084 (2014).
- 119 Kendzerska, T., Gershon, A. S., Hawker, G., Tomlinson, G. & Leung, R. S. Obstructive Sleep Apnea and Incident Diabetes: A Historical Cohort Study. *American journal of respiratory and critical care medicine*, doi:10.1164/rccm.201312-2209OC (2014).
- 120 Gan, Y. *et al.* Shift work and diabetes mellitus: a meta-analysis of observational studies. *Occupational and environmental medicine*, doi:10.1136/oemed-2014-102150 (2014).
- 121 Lind, L., Zethelius, B., Salihovic, S., van Bavel, B. & Lind, P. M. Circulating levels of perfluoroalkyl substances and prevalent diabetes in the elderly. *Diabetologia* **57**, 473-479, doi:10.1007/s00125-013-3126-3 (2014).
- 122 Köbberling, J. & Tillil, H. Empirical risk figures for first degree relatives of non-insulin dependent diabetics. *The genetics of diabetes mellitus*, 201-209 (1982).
- 123 Groop, L. *et al.* Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. *Diabetes* **45**, 1585-1593 (1996).
- 124 Newman, B. *et al.* Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia* **30**, 763-768 (1987).
- 125 Kaprio, J. *et al.* Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* **35**, 1060-1067 (1992).
- 126 Poulsen, P. *et al.* Heritability of insulin secretion, peripheral and hepatic insulin action, and intracellular glucose partitioning in young and old Danish twins. *Diabetes* **54**, 275-283 (2005).

- 127 Weber-Lehmann, J. *et al.* Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic science international. Genetics* **9**, 42-46, doi:10.1016/j.fsigen.2013.10.015 (2014).
- 128 Ounissi-Benkhalha, H. & Polychronakos, C. The molecular genetics of type 1 diabetes: new genes and emerging mechanisms. *Trends Mol Med* **14**, 268-275, doi:10.1016/j.molmed.2008.04.002 (2008).
- 129 Grarup, N., Sandholt, C. H., Hansen, T. & Pedersen, O. Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia* **57**, 1528-1541, doi:10.1007/s00125-014-3270-4 (2014).
- 130 Cauchi, S. *et al.* TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *Journal of molecular medicine* **85**, 777-782, doi:10.1007/s00109-007-0203-4 (2007).
- 131 Yamauchi, T. *et al.* A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* **42**, 864-868, doi:10.1038/ng.660 (2010).
- 132 Unoki, H. *et al.* SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* **40**, 1098-1102, doi:10.1038/ng.208 (2008).
- 133 Yasuda, K. *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* **40**, 1092-1097, doi:10.1038/ng.207 (2008).
- 134 Kooner, J. S. *et al.* Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* **43**, 984-989, doi:10.1038/ng.921 (2011).
- 135 Cho, Y. S. *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* **44**, 67-72, doi:10.1038/ng.1019 (2012).
- 136 Waters, K. M. *et al.* Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS Genet* **6**, doi:10.1371/journal.pgen.1001078 (2010).
- 137 Haiman, C. A. *et al.* Consistent directions of effect for established type 2 diabetes risk variants across populations: the population architecture using Genomics and Epidemiology (PAGE) Consortium. *Diabetes* **61**, 1642-1647, doi:10.2337/db11-1296 (2012).
- 138 Replication, D. I. G. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-244, doi:10.1038/ng.2897 (2014).
- 139 Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature advance online publication*, doi:10.1038/nature13425 (2014).
- 140 Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-990, doi:10.1038/ng.2383 (2012).
- 141 Tukiainen, T. *et al.* Chromosome X-wide association study identifies Loci for fasting insulin and height and evidence for incomplete dosage compensation. *PLoS Genet* **10**, e1004127, doi:10.1371/journal.pgen.1004127 (2014).
- 142 Kanakatti Shankar, R. *et al.* Permanent neonatal diabetes mellitus: prevalence and genetic diagnosis in the SEARCH for Diabetes in Youth Study. *Pediatric diabetes* **14**, 174-180, doi:10.1111/pedi.12003 (2013).
- 143 Rubio-Cabezas, O. *et al.* Wolcott-Rallison syndrome is the most common genetic cause of permanent neonatal diabetes in consanguineous families. *J Clin Endocrinol Metab* **94**, 4162-4170, doi:10.1210/jc.2009-1137 (2009).
- 144 Shaw-Smith, C. *et al.* GATA4 mutations are a cause of neonatal and childhood-onset diabetes. *Diabetes*, doi:10.2337/db14-0061 (2014).
- 145 Lango Allen, H. *et al.* GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet* **44**, 20-22, doi:10.1038/ng.1035 (2012).

- 146 Stoffers, D. A., Zinkin, N. T., Stanojevic, V., Clarke, W. L. & Habener, J. F. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat Genet* **15**, 106-110, doi:10.1038/ng0197-106 (1997).
- 147 Sellick, G. S. *et al.* Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat Genet* **36**, 1301-1305, doi:10.1038/ng1475 (2004).
- 148 Smith, S. B. *et al.* Rfx6 directs islet formation and insulin production in mice and humans. *Nature* **463**, 775-780, doi:10.1038/nature08748 (2010).
- 149 De Leon, D. D. & Stanley, C. A. in *GeneReviews* (eds R. A. Pagon *et al.*) (1993).
- 150 Thomas, P. M. *et al.* Mutations in the sulfonylurea receptor gene in familial persistent hyperinsulinemic hypoglycemia of infancy. *Science* **268**, 426-429 (1995).
- 151 Stanley, C. A. *et al.* Hyperinsulinism and hyperammonemia in infants with regulatory mutations of the glutamate dehydrogenase gene. *N Engl J Med* **338**, 1352-1357, doi:10.1056/NEJM199805073381904 (1998).
- 152 Christesen, H. B. *et al.* Activating glucokinase (GCK) mutations as a cause of medically responsive congenital hyperinsulinism: prevalence in children and characterisation of a novel GCK mutation. *European journal of endocrinology / European Federation of Endocrine Societies* **159**, 27-34, doi:10.1530/EJE-08-0203 (2008).
- 153 Clayton, P. T. *et al.* Hyperinsulinism in short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency reveals the importance of beta-oxidation in insulin secretion. *J Clin Invest* **108**, 457-465, doi:10.1172/JCI11294 (2001).
- 154 Stanescu, D. E., Hughes, N., Kaplan, B., Stanley, C. A. & De Leon, D. D. Novel presentations of congenital hyperinsulinism due to mutations in the MODY genes: HNF1A and HNF4A. *J Clin Endocrinol Metab* **97**, E2026-2030, doi:10.1210/jc.2012-1356 (2012).
- 155 Otonkoski, T. *et al.* Physical exercise-induced hypoglycemia caused by failed silencing of monocarboxylate transporter 1 in pancreatic beta cells. *Am J Hum Genet* **81**, 467-474, doi:10.1086/520960 (2007).
- 156 Gonzalez-Barroso, M. M. *et al.* Mutations in UCP2 in congenital hyperinsulinism reveal a role for regulation of insulin secretion. *PLoS One* **3**, e3850, doi:10.1371/journal.pone.0003850 (2008).
- 157 Polak, M. & Cave, H. Neonatal diabetes mellitus: a disease linked to multiple mechanisms. *Orphanet journal of rare diseases* **2**, 12, doi:10.1186/1750-1172-2-12 (2007).
- 158 Temple, I. K. & Shield, J. P. Transient neonatal diabetes, a disorder of imprinting. *Journal of medical genetics* **39**, 872-875 (2002).
- 159 Mackay, D. J. *et al.* Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet* **40**, 949-951, doi:10.1038/ng.187 (2008).
- 160 Boonen, S. E. *et al.* Transient neonatal diabetes, ZFP57, and hypomethylation of multiple imprinted loci: a detailed follow-up. *Diabetes Care* **36**, 505-512, doi:10.2337/dc12-0700 (2013).
- 161 Delepine, M. *et al.* EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. *Nat Genet* **25**, 406-409, doi:10.1038/78085 (2000).
- 162 Vionnet, N. *et al.* Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 721-722, doi:10.1038/356721a0 (1992).
- 163 Senee, V. *et al.* Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet* **38**, 682-687, doi:10.1038/ng1802 (2006).
- 164 Poulton, C. J. *et al.* Microcephaly with simplified gyration, epilepsy, and infantile diabetes linked to inappropriate apoptosis of neural progenitors. *Am J Hum Genet* **89**, 265-276, doi:10.1016/j.ajhg.2011.07.006 (2011).

- 165 Kwok, S. C., Chan, S. J., Rubenstein, A. H., Poucher, R. & Steiner, D. F. Loss of a restriction endonuclease cleavage site in the gene of a structurally abnormal human insulin. *Biochem Biophys Res Commun* **98**, 844-849 (1981).
- 166 Thomas, P., Ye, Y. & Lightner, E. Mutation of the pancreatic islet inward rectifier Kir6.2 also leads to familial persistent hyperinsulinemic hypoglycemia of infancy. *Hum Mol Genet* **5**, 1809-1812 (1996).
- 167 Bonnefond, A. *et al.* Transcription factor gene MNX1 is a novel cause of permanent neonatal diabetes in a consanguineous family. *Diabetes Metab* **39**, 276-280, doi:10.1016/j.diabet.2013.02.007 (2013).
- 168 Malecki, M. T. *et al.* Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat Genet* **23**, 323-328, doi:10.1038/15500 (1999).
- 169 Milord, E. & Gragnoli, C. NEUROG3 variants and type 2 diabetes in Italians. *Minerva medica* **97**, 373-378 (2006).
- 170 Flanagan, S. E. *et al.* Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. *Cell Metab* **19**, 146-154, doi:10.1016/j.cmet.2013.11.021 (2014).
- 171 Solomon, B. D. *et al.* Compound heterozygosity for mutations in PAX6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microphthalmia. *American journal of medical genetics. Part A* **149A**, 2543-2546, doi:10.1002/ajmg.a.33081 (2009).
- 172 Simaite, D. *et al.* Recessive mutations in PCBD1 cause a new type of early-onset diabetes. *Diabetes*, doi:10.2337/db13-1784 (2014).
- 173 Shaw-Smith, C. *et al.* Recessive SLC19A2 mutations are a cause of neonatal diabetes mellitus in thiamine-responsive megaloblastic anaemia. *Pediatric diabetes* **13**, 314-321, doi:10.1111/j.1399-5448.2012.00855.x (2012).
- 174 Tanizawa, Y. *et al.* Variability of the pancreatic islet beta cell/liver (GLUT 2) glucose transporter gene in NIDDM patients. *Diabetologia* **37**, 420-427 (1994).
- 175 Inoue, H. *et al.* A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat Genet* **20**, 143-148, doi:10.1038/2441 (1998).
- 176 Tattersall, R. B. & Fajans, S. S. A difference between the inheritance of classical juvenile-onset and maturity-onset type diabetes of young people. *Diabetes* **24**, 44-53 (1975).
- 177 Hattersley, A. T. *et al.* Linkage of type 2 diabetes to the glucokinase gene. *Lancet* **339**, 1307-1310 (1992).
- 178 Froguel, P. *et al.* Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 162-164, doi:10.1038/356162a0 (1992).
- 179 Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455-458, doi:10.1038/384455a0 (1996).
- 180 Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458-460, doi:10.1038/384458a0 (1996).
- 181 Stoffers, D. A., Ferrer, J., Clarke, W. L. & Habener, J. F. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nat Genet* **17**, 138-139, doi:10.1038/ng1097-138 (1997).
- 182 Horikawa, Y. *et al.* Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet* **17**, 384-385, doi:10.1038/ng1297-384 (1997).
- 183 Tanaka, D. *et al.* Exome sequencing identifies a new candidate mutation for susceptibility to diabetes in a family with highly aggregated type 2 diabetes. *Mol Genet Metab* **109**, 112-117, doi:10.1016/j.ymgme.2013.02.010 (2013).
- 184 Zhang, C. *et al.* MafA is a key regulator of glucose-stimulated insulin secretion. *Molecular and cellular biology* **25**, 4969-4976, doi:10.1128/MCB.25.12.4969-4976.2005 (2005).
- 185 Edghill, E. L. *et al.* Sequencing of candidate genes selected by beta cell experts in monogenic diabetes of unknown aetiology. *JOP : Journal of the pancreas* **11**, 14-17 (2010).

186 Igoillo-Esteve, M. *et al.* tRNA methyltransferase homolog gene TRMT10A mutation in young  
onset diabetes and primary microcephaly in humans. *PLoS Genet* **9**, e1003888,  
doi:10.1371/journal.pgen.1003888 (2013).

187 Prudente, S. in *EASD* (Vienna, Austria, 2014).

188 Johansson, S. *et al.* Exome sequencing and genetic testing for MODY. *PLoS One* **7**, e38050,  
doi:10.1371/journal.pone.0038050 (2012).

189 Reardon, W. *et al.* Diabetes mellitus associated with a pathogenic point mutation in  
mitochondrial DNA. *Lancet* **340**, 1376-1379 (1992).

190 Kameoka, K. *et al.* Novel mitochondrial DNA mutation in tRNA(Lys) (8296A-->G) associated  
with diabetes. *Biochem Biophys Res Commun* **245**, 523-527, doi:10.1006/bbrc.1998.8437  
(1998).

191 Hanna, M. G. *et al.* Congenital encephalomyopathy and adult-onset myopathy and diabetes  
mellitus: different phenotypic associations of a new heteroplasmic mtDNA tRNA glutamic  
acid mutation. *Am J Hum Genet* **56**, 1026-1033 (1995).

192 Bannwarth, S. *et al.* A novel unstable mutation in mitochondrial DNA responsible for  
maternally inherited diabetes and deafness. *Diabetes Care* **34**, 2591-2593, doi:10.2337/dc11-  
1012 (2011).

193 Neve, B. *et al.* Role of transcription factor KLF11 and its diabetes-associated gene variants in  
pancreatic beta cell function. *Proc Natl Acad Sci U S A* **102**, 4807-4812,  
doi:10.1073/pnas.0409177102 (2005).

194 Raeder, H. *et al.* Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic  
exocrine dysfunction. *Nat Genet* **38**, 54-62, doi:10.1038/ng1708 (2006).

195 Shimajiri, Y. *et al.* A missense mutation of Pax4 gene (R121W) is associated with type 2  
diabetes in Japanese. *Diabetes* **50**, 2864-2869 (2001).

196 Borowiec, M. *et al.* Mutations at the BLK locus linked to maturity onset diabetes of the young  
and beta-cell dysfunction. *Proc Natl Acad Sci U S A* **106**, 14460-14465,  
doi:10.1073/pnas.0906474106 (2009).

197 Bonnefond, A. *et al.* Whole-exome sequencing and high throughput genotyping identified  
KCNJ11 as the thirteenth MODY gene. *PLoS One* **7**, e37423,  
doi:10.1371/journal.pone.0037423 (2012).

198 Bonnycastle, L. L. *et al.* Common variants in maturity-onset diabetes of the young genes  
contribute to risk of type 2 diabetes in Finns. *Diabetes* **55**, 2534-2540, doi:10.2337/db06-  
0178 (2006).

199 S, B. *et al.* Vol. 29 OC3.1 (ed Endocrine Abstracts) (2012).

200 Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-  
associated variants. *Nat Genet* **46**, 136-143, doi:10.1038/ng.2870 (2014).

201 Weir, G. C. & Bonner-Weir, S. Five stages of evolving beta-cell dysfunction during progression  
to diabetes. *Diabetes* **53 Suppl 3**, S16-21 (2004).

202 Dimas, A. S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic  
traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158-2171, doi:10.2337/db13-0949  
(2014).

203 Frayling, T. M. & Hattersley, A. T. Physiology helps GWAS take a step closer to mechanism.  
*Diabetes* **63**, 1836-1837, doi:10.2337/db14-0130 (2014).

204 Matschinsky, F. M. Assessing the potential of glucokinase activators in diabetes therapy.  
*Nature reviews. Drug discovery* **8**, 399-416, doi:10.1038/nrd2850 (2009).

205 Muoio, D. M. & Newgard, C. B. Mechanisms of disease: molecular and metabolic mechanisms  
of insulin resistance and beta-cell failure in type 2 diabetes. *Nature reviews. Molecular cell  
biology* **9**, 193-205, doi:10.1038/nrm2327 (2008).

206 Koster, J. C., Permutt, M. A. & Nichols, C. G. Diabetes and insulin secretion: the ATP-sensitive  
K<sup>+</sup> channel (K ATP) connection. *Diabetes* **54**, 3065-3072 (2005).

207 Rajan, A. S. *et al.* Ion channels and insulin secretion. *Diabetes Care* **13**, 340-363 (1990).



- 208 Fajans, S. S., Bell, G. I. & Polonsky, K. S. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* **345**, 971-980, doi:10.1056/NEJMra002168 (2001).
- 209 Karges, B., Meissner, T., Icks, A., Kapellen, T. & Holl, R. W. Management of diabetes mellitus in infants. *Nat Rev Endocrinol* **8**, 201-211, doi:10.1038/nrendo.2011.204 (2012).
- 210 Michau, A. *et al.* Mutations in SLC2A2 gene reveal hGLUT2 function in pancreatic beta cell development. *J Biol Chem* **288**, 31080-31092, doi:10.1074/jbc.M113.469189 (2013).
- 211 Rothman, D. L. *et al.* Decreased muscle glucose transport/phosphorylation is an early defect in the pathogenesis of non-insulin-dependent diabetes mellitus. *Proc Natl Acad Sci U S A* **92**, 983-987 (1995).
- 212 Basu, A. *et al.* Effects of type 2 diabetes on the ability of insulin and glucose to regulate splanchnic and muscle glucose metabolism: evidence for a defect in hepatic glucokinase activity. *Diabetes* **49**, 272-283 (2000).
- 213 Proks, P., Girard, C. & Ashcroft, F. M. Functional effects of KCNJ11 mutations causing neonatal diabetes: enhanced activation by MgATP. *Hum Mol Genet* **14**, 2717-2726, doi:10.1093/hmg/ddi305 (2005).
- 214 Vaxillaire, M. *et al.* New ABCC8 mutations in relapsing neonatal diabetes and clinical features. *Diabetes* **56**, 1737-1741, doi:10.2337/db06-1540 (2007).
- 215 Colombo, C. *et al.* Seven mutations in the human insulin gene linked to permanent neonatal/infancy-onset diabetes mellitus. *J Clin Invest* **118**, 2148-2156, doi:10.1172/JCI33777 (2008).
- 216 Yamada, T. *et al.* WFS1-deficiency increases endoplasmic reticulum stress, impairs cell cycle progression and triggers the apoptotic pathway specifically in pancreatic beta-cells. *Hum Mol Genet* **15**, 1600-1609, doi:10.1093/hmg/ddl081 (2006).
- 217 Erdmann, S. *et al.* Tissue-specific transcription factor HNF4alpha inhibits cell proliferation and induces apoptosis in the pancreatic INS-1 beta-cell line. *Biological chemistry* **388**, 91-106, doi:10.1515/BC.2007.011 (2007).
- 218 Brickwood, S. *et al.* Wolcott-Rallison syndrome: pathogenic insights into neonatal diabetes from new mutation and expression studies of EIF2AK3. *Journal of medical genetics* **40**, 685-689 (2003).
- 219 Haldorsen, I. S. *et al.* Lack of pancreatic body and tail in HNF1B mutation carriers. *Diabetic medicine : a journal of the British Diabetic Association* **25**, 782-787, doi:10.1111/j.1464-5491.2008.02460.x (2008).
- 220 Oliver-Krasinski, J. M. *et al.* The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice. *J Clin Invest* **119**, 1888-1898, doi:10.1172/JCI37028 (2009).
- 221 Krapp, A. *et al.* The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. *Genes Dev* **12**, 3752-3763 (1998).
- 222 Soyer, J. *et al.* Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. *Development* **137**, 203-212, doi:10.1242/dev.041673 (2010).
- 223 Kang, H. S. *et al.* Transcription factor Glis3, a novel critical player in the regulation of pancreatic beta-cell development and insulin gene expression. *Molecular and cellular biology* **29**, 6366-6379, doi:10.1128/MCB.01259-09 (2009).
- 224 St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A. & Gruss, P. Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. *Nature* **387**, 406-409, doi:10.1038/387406a0 (1997).
- 225 Itkin-Ansari, P. *et al.* NeuroD1 in the endocrine pancreas: localization and dual function as an activator and repressor. *Developmental dynamics : an official publication of the American Association of Anatomists* **233**, 946-953, doi:10.1002/dvdy.20443 (2005).

- 226 Wang, S. *et al.* Neurog3 gene dosage regulates allocation of endocrine and exocrine cell fates in the developing mouse pancreas. *Developmental biology* **339**, 26-37, doi:10.1016/j.ydbio.2009.12.009 (2010).
- 227 Servitja, J. M. *et al.* Hnf1alpha (MODY3) controls tissue-specific transcriptional programs and exerts opposed effects on cell growth in pancreatic islets and liver. *Molecular and cellular biology* **29**, 2945-2959, doi:10.1128/MCB.01389-08 (2009).
- 228 Miura, A. *et al.* Hepatocyte nuclear factor-4alpha is essential for glucose-stimulated insulin secretion by pancreatic beta-cells. *J Biol Chem* **281**, 5246-5257, doi:10.1074/jbc.M507496200 (2006).
- 229 Haslam, D. & Rigby, N. A long look at obesity. *The Lancet* **376**, 85-86 (2010).
- 230 Tipton, C. M. Susruta of India, an unrecognized contributor to the history of exercise physiology. *Journal of applied physiology* **104**, 1553-1556 (2008).
- 231 Bray, G. A. What's in a Name? Mr. Dickens' "Pickwickian" Fat Boy Syndrome. *Obesity research* **2**, 380-383, doi:10.1002/j.1550-8528.1994.tb00079.x (1994).
- 232 Chaun, H. Sir William Osler and gastroenterology. *Canadian journal of gastroenterology = Journal canadien de gastroenterologie* **24**, 615-618 (2010).
- 233 Bray, G. A. Life Insurance and Overweight. *Obesity research* **3**, 97-99, doi:10.1002/j.1550-8528.1995.tb00125.x (1995).
- 234 Neill, U. S. Leaping for leptin: the 2010 Albert Lasker Basic Medical Research Award goes to Douglas Coleman and Jeffrey M. Friedman. *J. Clin. Invest.* **120**, 3413-3418, doi:10.1172/JCI45094 (2010).
- 235 Association, A. M. AMA Adopts New Policies on Second Day of Voting at Annual Meeting, <<http://www.ama-assn.org/ama/pub/news/news/2013/2013-06-18-new-ama-policies-annual-meeting.page>> (2013).
- 236 Weaver, L. T. In the balance: weighing babies and the birth of the infant welfare clinic. *Bulletin of the history of medicine* **84**, 30-57 (2010).
- 237 Dublin LI, L. A. *Twenty-Five Years of Health Progress: A Study of the Mortality Experience Among Industrial Policyholders of the Metropolitan Life Insurance Company 1911-1935.* (Metropolitan Life Insurance Company, 1937).
- 238 Garrow, J. S. & Webster, J. Quetelet's index (W/H<sup>2</sup>) as a measure of fatness. *International journal of obesity* **9**, 147-153 (1985).
- 239 Eknayan, G. Adolphe Quetelet (1796-1874)--the average man and indices of obesity. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* **23**, 47-51, doi:10.1093/ndt/gfm517 (2008).
- 240 Okorodudu, D. O. *et al.* Diagnostic performance of body mass index to identify obesity as defined by body adiposity: a systematic review and meta-analysis. *Int J Obes (Lond)* **34**, 791-799, doi:10.1038/ijo.2010.5 (2010).
- 241 Freedman, D. S. *et al.* The body adiposity index (hip circumference / height(1.5)) is not a more accurate measure of adiposity than is BMI, waist circumference, or hip circumference. *Obesity* **20**, 2438-2444, doi:10.1038/oby.2012.81 (2012).
- 242 Tuan, N. T. & Wang, Y. Adiposity assessments: agreement between dual-energy X-ray absorptiometry and anthropometric measures in U.S. children. *Obesity* **22**, 1495-1504, doi:10.1002/oby.20689 (2014).
- 243 Ahima, R. S. & Lazar, M. A. Physiology. The health risk of obesity--better metrics imperative. *Science* **341**, 856-858, doi:10.1126/science.1241244 (2013).
- 244 Lubura, M. *et al.* Non-invasive quantification of white and brown adipose tissues and liver fat content by computed tomography in mice. *PLoS One* **7**, e37026, doi:10.1371/journal.pone.0037026 (2012).
- 245 Reddy, N. L. *et al.* Identification of brown adipose tissue using MR imaging in a human adult with histological and immunohistochemical confirmation. *J Clin Endocrinol Metab* **99**, E117-121, doi:10.1210/jc.2013-2036 (2014).



- 246 Rasmussen, J. M. *et al.* Brown adipose tissue quantification in human neonates using water-fat separated MRI. *PLoS One* **8**, e77907, doi:10.1371/journal.pone.0077907 (2013).
- 247 Organization, W. H. *Obesity and overweight*, <<http://www.who.int/mediacentre/factsheets/fs311/en/>> (2014).
- 248 Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, doi:10.1016/S0140-6736(14)60460-8 (2014).
- 249 Finucane, M. M. *et al.* National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet* **377**, 557-567, doi:10.1016/S0140-6736(10)62037-5 (2011).
- 250 Finkelstein, E. A. *et al.* Obesity and severe obesity forecasts through 2030. *American journal of preventive medicine* **42**, 563-570, doi:10.1016/j.amepre.2011.10.026 (2012).
- 251 Withrow, D. & Alter, D. A. The economic burden of obesity worldwide: a systematic review of the direct costs of obesity. *Obesity reviews : an official journal of the International Association for the Study of Obesity* **12**, 131-141, doi:10.1111/j.1467-789X.2009.00712.x (2011).
- 252 Wang, Y., Beydoun, M. A., Liang, L., Caballero, B. & Kumanyika, S. K. Will all Americans become overweight or obese? estimating the progression and cost of the US obesity epidemic. *Obesity* **16**, 2323-2330, doi:10.1038/oby.2008.351 (2008).
- 253 ROCHE, I. K. H. ObEpi-Roche 2012 : enquête nationale sur l'obésité et le surpoids. (Roche, 2012).
- 254 Emery, C. *et al.* [Cost of obesity in France]. *Presse medicale* **36**, 832-840, doi:10.1016/j.lpm.2006.12.031 (2007).
- 255 Kitahara, C. M. *et al.* Association between Class III Obesity (BMI of 40-59 kg/m<sup>2</sup>) and Mortality: A Pooled Analysis of 20 Prospective Studies. *PLoS medicine* **11**, e1001673, doi:10.1371/journal.pmed.1001673 (2014).
- 256 Mokdad, A. H. *et al.* The continuing epidemics of obesity and diabetes in the United States. *JAMA : the journal of the American Medical Association* **286**, 1195-1200 (2001).
- 257 Resnick, H. E., Valsania, P., Halter, J. B. & Lin, X. Relation of weight gain and weight loss on subsequent diabetes risk in overweight adults. *Journal of epidemiology and community health* **54**, 596-602 (2000).
- 258 Hillier, T. A. & Pedula, K. L. Characteristics of an adult population with newly diagnosed type 2 diabetes: the relation of obesity and age of onset. *Diabetes Care* **24**, 1522-1527 (2001).
- 259 Balkau, B. *et al.* International Day for the Evaluation of Abdominal Obesity (IDEA): a study of waist circumference, cardiovascular disease, and diabetes mellitus in 168,000 primary care patients in 63 countries. *Circulation* **116**, 1942-1951, doi:10.1161/CIRCULATIONAHA.106.676379 (2007).
- 260 Boyko, E. J., Fujimoto, W. Y., Leonetti, D. L. & Newell-Morris, L. Visceral adiposity and risk of type 2 diabetes: a prospective study among Japanese Americans. *Diabetes Care* **23**, 465-471 (2000).
- 261 Ding, J. *et al.* The association of regional fat depots with hypertension in older persons of white and African American ethnicity. *American journal of hypertension* **17**, 971-976, doi:10.1016/j.amjhyper.2004.05.001 (2004).
- 262 Hayashi, T. *et al.* Visceral adiposity is an independent predictor of incident hypertension in Japanese Americans. *Annals of internal medicine* **140**, 992-1000 (2004).
- 263 Must, A. *et al.* The disease burden associated with overweight and obesity. *JAMA : the journal of the American Medical Association* **282**, 1523-1529 (1999).
- 264 Vasan, R. S., Larson, M. G., Leip, E. P., Kannel, W. B. & Levy, D. Assessment of frequency of progression to hypertension in non-hypertensive participants in the Framingham Heart Study: a cohort study. *Lancet* **358**, 1682-1686, doi:10.1016/S0140-6736(01)06710-1 (2001).

- 265 Hokanson, J. E. & Austin, M. A. Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. *Journal of cardiovascular risk* **3**, 213-219 (1996).
- 266 Yuan, G., Al-Shali, K. Z. & Hegele, R. A. Hypertriglyceridemia: its etiology, effects and treatment. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* **176**, 1113-1120, doi:10.1503/cmaj.060963 (2007).
- 267 Taskinen, M. R. Diabetic dyslipidaemia: from basic research to clinical practice. *Diabetologia* **46**, 733-749, doi:10.1007/s00125-003-1111-y (2003).
- 268 Grundy, S. M. Hypertriglyceridemia, atherogenic dyslipidemia, and the metabolic syndrome. *The American journal of cardiology* **81**, 18B-25B (1998).
- 269 Alberti, K. G. *et al.* Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* **120**, 1640-1645, doi:10.1161/CIRCULATIONAHA.109.192644 (2009).
- 270 Ho, J. S. *et al.* Relation of the number of metabolic syndrome risk factors with all-cause and cardiovascular mortality. *The American journal of cardiology* **102**, 689-692, doi:10.1016/j.amjcard.2008.05.010 (2008).
- 271 Thomsen, M. & Nordestgaard, B. G. Myocardial infarction and ischemic heart disease in overweight and obesity with and without metabolic syndrome. *JAMA internal medicine* **174**, 15-22, doi:10.1001/jamainternmed.2013.10522 (2014).
- 272 Despres, J. P. & Lemieux, I. Abdominal obesity and metabolic syndrome. *Nature* **444**, 881-887, doi:10.1038/nature05488 (2006).
- 273 Bahrami, H. *et al.* Novel metabolic risk factors for incident heart failure and their relationship with obesity: the MESA (Multi-Ethnic Study of Atherosclerosis) study. *Journal of the American College of Cardiology* **51**, 1775-1783, doi:10.1016/j.jacc.2007.12.048 (2008).
- 274 Goodpaster, B. H. *et al.* Obesity, regional body fat distribution, and the metabolic syndrome in older men and women. *Arch Intern Med* **165**, 777-783, doi:10.1001/archinte.165.7.777 (2005).
- 275 Yatsuya, H. *et al.* Race- and sex-specific associations of obesity measures with ischemic stroke incidence in the Atherosclerosis Risk in Communities (ARIC) study. *Stroke; a journal of cerebral circulation* **41**, 417-425, doi:10.1161/STROKEAHA.109.566299 (2010).
- 276 Arnold, M. *et al.* Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *The Lancet Oncology*, doi:[http://dx.doi.org/10.1016/S1470-2045\(14\)71123-4](http://dx.doi.org/10.1016/S1470-2045(14)71123-4).
- 277 Calle, E. E., Rodriguez, C., Walker-Thurmond, K. & Thun, M. J. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med* **348**, 1625-1638, doi:10.1056/NEJMoa021423 (2003).
- 278 Basen-Engquist, K. & Chang, M. Obesity and cancer risk: recent review and evidence. *Current oncology reports* **13**, 71-76, doi:10.1007/s11912-010-0139-7 (2011).
- 279 Everhart, J. E. Contributions of obesity and weight loss to gallstone disease. *Annals of internal medicine* **119**, 1029-1035 (1993).
- 280 Friedman, G. D., Kannel, W. B. & Dawber, T. R. The epidemiology of gallbladder disease: observations in the Framingham Study. *Journal of chronic diseases* **19**, 273-292 (1966).
- 281 Felson, D. T. Weight and osteoarthritis. *The Journal of rheumatology. Supplement* **43**, 7-9 (1995).
- 282 Pi-Sunyer, F. X. Medical hazards of obesity. *Annals of internal medicine* **119**, 655-660 (1993).
- 283 Luppino, F. S. *et al.* Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Archives of general psychiatry* **67**, 220-229, doi:10.1001/archgenpsychiatry.2010.2 (2010).

284 Helling, T. S. & Gurram, K. Nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and  
 bariatric surgery: a review. *Surgery for obesity and related diseases : official journal of the  
 American Society for Bariatric Surgery* **2**, 213-220, doi:10.1016/j.soard.2006.01.002 (2006).

285 Klein, S. Is visceral fat responsible for the metabolic abnormalities associated with obesity?:  
 implications of omentectomy. *Diabetes Care* **33**, 1693-1694, doi:10.2337/dc10-0744 (2010).

286 Lowell, B. B. *et al.* Development of obesity in transgenic mice after genetic ablation of brown  
 adipose tissue. *Nature* **366**, 740-742, doi:10.1038/366740a0 (1993).

287 Ravussin, E., Lillioja, S., Anderson, T. E., Christin, L. & Bogardus, C. Determinants of 24-hour  
 energy expenditure in man. Methods and results using a respiratory chamber. *J Clin Invest*  
**78**, 1568-1578, doi:10.1172/JCI112749 (1986).

288 B, B. *Intracranial Tumours*. (1888).

289 A, F. Ein fall von tumor der hypophysis cerebri ohne akromegalie. 883 (1901).

290 Hetherington, A. W. & Ranson, S. W. Hypothalamic lesions and adiposity in the rat. *The  
 Anatomical Record* **78**, 149-172, doi:10.1002/ar.1090780203 (1940).

291 Ingalls, A. M., Dickie, M. M. & Snell, G. D. Obese, a new mutation in the house mouse. *The  
 Journal of heredity* **41**, 317-318 (1950).

292 Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632-  
 635, doi:10.1038/379632a0 (1996).

293 Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**,  
 1263-1271 (1995).

294 Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue.  
*Nature* **372**, 425-432, doi:10.1038/372425a0 (1994).

295 Campfield, L. A., Smith, F. J., Guisez, Y., Devos, R. & Burn, P. Recombinant mouse OB protein:  
 evidence for a peripheral signal linking adiposity and central neural networks. *Science* **269**,  
 546-549 (1995).

296 Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the obese gene.  
*Science* **269**, 543-546 (1995).

297 Halaas, J. L. *et al.* Physiological response to long-term peripheral and central leptin infusion in  
 lean and obese mice. *Proc Natl Acad Sci U S A* **94**, 8878-8883 (1997).

298 Pellemounter, M. A. *et al.* Effects of the obese gene product on body weight regulation in  
 ob/ob mice. *Science* **269**, 540-543 (1995).

299 Miller, M. W. *et al.* Cloning of the mouse agouti gene predicts a secreted protein ubiquitously  
 expressed in mice carrying the lethal yellow mutation. *Genes Dev* **7**, 454-467 (1993).

300 Bultman, S. J., Michaud, E. J. & Woychik, R. P. Molecular characterization of the mouse agouti  
 locus. *Cell* **71**, 1195-1204 (1992).

301 Lu, D. *et al.* Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor.  
*Nature* **371**, 799-802, doi:10.1038/371799a0 (1994).

302 Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic  
 neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165-168,  
 doi:10.1038/385165a0 (1997).

303 Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in  
 mice. *Cell* **88**, 131-141 (1997).

304 Melmed, S. & Williams, R. H. *Williams textbook of endocrinology*. 12th ed. / Shlomo Melmed  
 ... [et al.] edn, (Philadelphia : Elsevier/Saunders, c2011).

305 Kennedy, G. C. The role of depot fat in the hypothalamic control of food intake in the rat.  
*Proceedings of the Royal Society of London. Series B, Containing papers of a Biological  
 character. Royal Society* **140**, 578-596 (1953).

306 Srivastava, N., Lakhani, R. & Mittal, B. Pathophysiology and genetics of obesity. *Indian journal  
 of experimental biology* **45**, 929-936 (2007).

307 Marty, N., Dallaporta, M. & Thorens, B. Brain glucose sensing, counterregulation, and energy  
 homeostasis. *Physiology* **22**, 241-251, doi:10.1152/physiol.00010.2007 (2007).

- 308 Levin, B. E., Kang, L., Sanders, N. M. & Dunn-Meynell, A. A. Role of Neuronal Glucosensing in the Regulation of Energy Homeostasis. *Diabetes* **55**, S122-S130, doi:10.2337/db06-S016 (2006).
- 309 Mobbs, C. V. *et al.* Impaired glucose signaling as a cause of obesity and the metabolic syndrome: the glucoadipostatic hypothesis. *Physiology & behavior* **85**, 3-23, doi:10.1016/j.physbeh.2005.04.005 (2005).
- 310 Brobeck, J. R. Food intake as a mechanism of temperature regulation. *The Yale journal of biology and medicine* **20**, 545-552 (1948).
- 311 Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report. National Institutes of Health. *Obesity research* **6 Suppl 2**, 51S-209S (1998).
- 312 KD, B. The LEARN Program for Weight Management. (Dallas, TX, 2000).
- 313 Program, D. P. Diabetes Prevention Program Lifestyle Core. (1996).
- 314 Wadden, T. A. & Butryn, M. L. Behavioral treatment of obesity. *Endocrinology and metabolism clinics of North America* **32**, 981-1003, x (2003).
- 315 Jeffery, R. W. *et al.* Long-term maintenance of weight loss: current status. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association* **19**, 5-16 (2000).
- 316 Wadden, T. A., Butryn, M. L. & Byrne, K. J. Efficacy of lifestyle modification for long-term weight control. *Obesity research* **12 Suppl**, 151S-162S, doi:10.1038/oby.2004.282 (2004).
- 317 Bray GA, B. C. *Handbook of obesity: Clinical applications*. 185-199 (Marcel Dekker, Inc).
- 318 Anderson, J. W., Konz, E. C., Frederich, R. C. & Wood, C. L. Long-term weight-loss maintenance: a meta-analysis of US studies. *Am J Clin Nutr* **74**, 579-584 (2001).
- 319 Look, A. R. G. & Wing, R. R. Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the Look AHEAD trial. *Arch Intern Med* **170**, 1566-1575, doi:10.1001/archinternmed.2010.334 (2010).
- 320 Nordmann, A. J. *et al.* Effects of low-carbohydrate vs low-fat diets on weight loss and cardiovascular risk factors: a meta-analysis of randomized controlled trials. *Arch Intern Med* **166**, 285-293, doi:10.1001/archinte.166.3.285 (2006).
- 321 Greenberg, I., Stampfer, M. J., Schwarzfuchs, D., Shai, I. & Group, D. Adherence and success in long-term weight loss diets: the dietary intervention randomized controlled trial (DIRECT). *Journal of the American College of Nutrition* **28**, 159-168 (2009).
- 322 Maclean, P. S., Bergouignan, A., Cornier, M. A. & Jackman, M. R. Biology's response to dieting: the impetus for weight regain. *American journal of physiology. Regulatory, integrative and comparative physiology* **301**, R581-600, doi:10.1152/ajpregu.00755.2010 (2011).
- 323 Burns, E. M. *et al.* Introduction of laparoscopic bariatric surgery in England: observational population cohort study. *BMJ* **341**, c4296, doi:10.1136/bmj.c4296 (2010).
- 324 Ryan, K. K. *et al.* FXR is a molecular target for the effects of vertical sleeve gastrectomy. *Nature* **509**, 183-188, doi:10.1038/nature13135 (2014).
- 325 Miras, A. D. & le Roux, C. W. Mechanisms underlying weight loss after bariatric surgery. *Nature reviews. Gastroenterology & hepatology* **10**, 575-584, doi:10.1038/nrgastro.2013.119 (2013).
- 326 Brethauer, S. A., Hammel, J. P. & Schauer, P. R. Systematic review of sleeve gastrectomy as staging and primary bariatric procedure. *Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery* **5**, 469-475, doi:10.1016/j.soard.2009.05.011 (2009).
- 327 O'Brien, P. E., McPhail, T., Chaston, T. B. & Dixon, J. B. Systematic review of medium-term weight loss after bariatric operations. *Obesity surgery* **16**, 1032-1040, doi:10.1381/096089206778026316 (2006).

- 328 Puzziferri, N. *et al.* Long-term follow-up after bariatric surgery: a systematic review. *JAMA : the journal of the American Medical Association* **312**, 934-942, doi:10.1001/jama.2014.10706 (2014).
- 329 Casagrande, D. S. *et al.* Incidence of Cancer Following Bariatric Surgery: Systematic Review and Meta-analysis. *Obesity surgery*, doi:10.1007/s11695-014-1276-0 (2014).
- 330 Tee, M. C., Cao, Y., Warnock, G. L., Hu, F. B. & Chavarro, J. E. Effect of bariatric surgery on oncologic outcomes: a systematic review and meta-analysis. *Surgical endoscopy* **27**, 4449-4456, doi:10.1007/s00464-013-3127-9 (2013).
- 331 Buchwald, H. *et al.* Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. *The American journal of medicine* **122**, 248-256 e245, doi:10.1016/j.amjmed.2008.09.041 (2009).
- 332 Pontiroli, A. E. & Morabito, A. Long-term prevention of mortality in morbid obesity through bariatric surgery. a systematic review and meta-analysis of trials performed with gastric banding and gastric bypass. *Annals of surgery* **253**, 484-487, doi:10.1097/SLA.0b013e31820d98cb (2011).
- 333 Vest, A. R., Heneghan, H. M., Agarwal, S., Schauer, P. R. & Young, J. B. Bariatric surgery and cardiovascular outcomes: a systematic review. *Heart* **98**, 1763-1777, doi:10.1136/heartjnl-2012-301778 (2012).
- 334 Schauer, P. R. *et al.* Bariatric Surgery versus Intensive Medical Therapy for Diabetes — 3-Year Outcomes. *New England Journal of Medicine* **0**, null, doi:doi:10.1056/NEJMoa1401329.
- 335 Carlsson, L. M. *et al.* Bariatric surgery and prevention of type 2 diabetes in Swedish obese subjects. *N Engl J Med* **367**, 695-704, doi:10.1056/NEJMoa1112082 (2012).
- 336 Ivezaj, V. & Grilo, C. M. When Mood Worsens after Gastric Bypass Surgery: Characterization of Bariatric Patients with Increases in Depressive Symptoms Following Surgery. *Obesity surgery*, doi:10.1007/s11695-014-1402-z (2014).
- 337 Schievink, W. I., Goseland, A. & Cunneen, S. Bariatric surgery as a possible risk factor for spontaneous intracranial hypotension. *Neurology*, doi:10.1212/WNL.0000000000000985 (2014).
- 338 Adams, T. D. *et al.* Long-term mortality after gastric bypass surgery. *N Engl J Med* **357**, 753-761, doi:10.1056/NEJMoa066603 (2007).
- 339 Hemmingsson, E., Johansson, K. & Reynisdottir, S. Effects of childhood abuse on adult obesity: a systematic review and meta-analysis. *Obesity reviews : an official journal of the International Association for the Study of Obesity*, doi:10.1111/obr.12216 (2014).
- 340 Fuller-Thomson, E., Sinclair, D. A. & Brennenstuhl, S. Carrying the pain of abuse: gender-specific findings on the relationship between childhood physical abuse and obesity in adulthood. *Obesity facts* **6**, 325-336, doi:10.1159/000354609 (2013).
- 341 Bhasin, S., Wallace, W., Lawrence, J. B. & Lesch, M. Sudden death associated with thyroid hormone abuse. *The American journal of medicine* **71**, 887-890 (1981).
- 342 COuncil on pharmacy and chemistry. *Journal of the American Medical Association* **105**, 31-33, doi:10.1001/jama.1935.02760270033012 (1935).
- 343 Asher, W. L. Mortality rate in patients receiving "diet pills". *Current therapeutic research, clinical and experimental* **14**, 525-539 (1972).
- 344 Connolly, H. M. *et al.* Valvular heart disease associated with fenfluramine-phentermine. *N Engl J Med* **337**, 581-588, doi:10.1056/NEJM199708283370901 (1997).
- 345 Kim, G. W., Lin, J. E., Blomain, E. S. & Waldman, S. A. New advances in models and strategies for developing anti-obesity drugs. *Expert opinion on drug discovery* **8**, 655-671, doi:10.1517/17460441.2013.792804 (2013).
- 346 Agency, E. M. *Withdrawal of the marketing authorisation application for Belviiq (lorcaserin)*, <[http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Medicine\\_QA/2013/05/WC500143811.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Medicine_QA/2013/05/WC500143811.pdf)> (2013).



- Agency, E. M. *Refusal of the marketing authorisation for Qsiva (phentermine/topiramate)*, <[http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Summary\\_of\\_opinion\\_-\\_Initial\\_authorisation/human/002350/WC500139215.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/002350/WC500139215.pdf)> (2013).
- Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214, doi:10.1126/science.1241214 (2013).
- Lane, J. A. *et al.* Randomised clinical trial: Helicobacter pylori eradication is associated with a significantly increased body mass index in a placebo-controlled study. *Alimentary pharmacology & therapeutics* **33**, 922-929, doi:10.1111/j.1365-2036.2011.04610.x (2011).
- Lender, N. *et al.* Review article: associations between Helicobacter pylori and obesity - an ecological study. *Alimentary pharmacology & therapeutics* **40**, 24-31, doi:10.1111/apt.12790 (2014).
- Francois, F. *et al.* The effect of H. pylori eradication on meal-associated changes in plasma ghrelin and leptin. *BMC gastroenterology* **11**, 37, doi:10.1186/1471-230X-11-37 (2011).
- Yang, Y. J., Sheu, B. S., Yang, H. B., Lu, C. C. & Chuang, C. C. Eradication of Helicobacter pylori increases childhood growth and serum acylated ghrelin levels. *World journal of gastroenterology : WJG* **18**, 2674-2681, doi:10.3748/wjg.v18.i21.2674 (2012).
- Buhling, A., Radun, D., Muller, W. A. & Malfertheiner, P. Influence of anti-Helicobacter triple-therapy with metronidazole, omeprazole and clarithromycin on intestinal microflora. *Alimentary pharmacology & therapeutics* **15**, 1445-1452 (2001).
- Heyman, L. *et al.* Evaluation of Beneficial Metabolic Effects of Berries in High-Fat Fed C57BL/6J Mice. *J Nutr Metab* **2014**, 403041, doi:10.1155/2014/403041 (2014).
- Chudnovskiy, R. *et al.* Consumption of Clarified Grapefruit Juice Ameliorates High-Fat Diet Induced Insulin Resistance and Weight Gain in Mice. *PLoS One* **9**, e108408, doi:10.1371/journal.pone.0108408 (2014).
- Morrissey, T. W., Jacknowitz, A. & Vinopal, K. Local food prices and their associations with children's weight and food security. *Pediatrics* **133**, 422-430, doi:10.1542/peds.2013-1963 (2014).
- Lobstein, T. Maternal and child obesity: some policy challenges. *The Proceedings of the Nutrition Society* **70**, 506-513, doi:10.1017/S0029665111003156 (2011).
- Ken, I. A Healthy Bottom Line: Obese Children, a Pacified Public, and Corporate Legitimacy. *Social Currents* **1**, 130-148, doi:10.1177/2329496514524927 (2014).
- Ezzati, M. *et al.* Rethinking the "diseases of affluence" paradigm: global patterns of nutritional risks in relation to economic development. *PLoS medicine* **2**, e133, doi:10.1371/journal.pmed.0020133 (2005).
- Monteiro, C. A., Moura, E. C., Conde, W. L. & Popkin, B. M. Socioeconomic status and obesity in adult populations of developing countries: a review. *Bulletin of the World Health Organization* **82**, 940-946, doi:/S0042-96862004001200011 (2004).
- World Urbanization Prospects: The 2011 Revision. (United Nations, Department of Economic and Social Affairs, Population Division, 2012).
- Ebrahim, S. *et al.* The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PLoS medicine* **7**, e1000268, doi:10.1371/journal.pmed.1000268 (2010).
- Villareal, D. T. *et al.* Weight loss, exercise, or both and physical function in obese older adults. *N Engl J Med* **364**, 1218-1229, doi:10.1056/NEJMoa1008234 (2011).
- Gilbert-Diamond, D., Li, Z., Adachi-Mejia, A. M., McClure, A. C. & Sargent, J. D. Association of a television in the bedroom with increased adiposity gain in a nationally representative sample of children and adolescents. *JAMA pediatrics* **168**, 427-434, doi:10.1001/jamapediatrics.2013.3921 (2014).
- Lear, S. A. *et al.* The association between ownership of common household devices and obesity and diabetes in high, middle and low income countries. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* **186**, 258-266, doi:10.1503/cmaj.131090 (2014).

- 366 Ladabaum, U., Mannalithara, A., Myer, P. A. & Singh, G. Obesity, Abdominal Obesity, Physical Activity, and Caloric Intake in U.S. Adults: 1988-2010. *The American journal of medicine*, doi:10.1016/j.amjmed.2014.02.026 (2014).
- 367 Hurley, B. F. *et al.* Muscle triglyceride utilization during exercise: effect of training. *Journal of applied physiology* **60**, 562-567 (1986).
- 368 Knutson, K. L. Does inadequate sleep play a role in vulnerability to obesity? *American journal of human biology : the official journal of the Human Biology Council* **24**, 361-371, doi:10.1002/ajhb.22219 (2012).
- 369 Spiegel, K., Leproult, R. & Van Cauter, E. Impact of sleep debt on metabolic and endocrine function. *Lancet* **354**, 1435-1439, doi:10.1016/S0140-6736(99)01376-8 (1999).
- 370 Taheri, S., Lin, L., Austin, D., Young, T. & Mignot, E. Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. *PLoS medicine* **1**, e62, doi:10.1371/journal.pmed.0010062 (2004).
- 371 Reid, K. J. *et al.* Timing and intensity of light correlate with body weight in adults. *PLoS One* **9**, e92251, doi:10.1371/journal.pone.0092251 (2014).
- 372 Johnson, F., Mavrogianni, A., Ucci, M., Vidal-Puig, A. & Wardle, J. Could increased time spent in a thermal comfort zone contribute to population increases in obesity? *Obesity reviews : an official journal of the International Association for the Study of Obesity* **12**, 543-551, doi:10.1111/j.1467-789X.2010.00851.x (2011).
- 373 Lee, P. *et al.* Temperature-acclimated brown adipose tissue modulates insulin sensitivity in humans. *Diabetes*, doi:10.2337/db14-0513 (2014).
- 374 Misra, A. & Khurana, L. Obesity and the metabolic syndrome in developing countries. *J Clin Endocrinol Metab* **93**, S9-30, doi:10.1210/jc.2008-1595 (2008).
- 375 De Vogli, R., Kouvonen, A. & Gimeno, D. 'Globesization': ecological evidence on the relationship between fast food outlets and obesity among 26 advanced economies. *Critical Public Health* **21**, 395-402, doi:10.1080/09581596.2011.619964 (2011).
- 376 Cetateanu, A. & Jones, A. Understanding the relationship between food environments, deprivation and childhood overweight and obesity: evidence from a cross sectional England-wide study. *Health & place* **27**, 68-76, doi:10.1016/j.healthplace.2014.01.007 (2014).
- 377 Miller, L. J., Joyce, S., Carter, S. & Yun, G. Associations Between Childhood Obesity and the Availability of Food Outlets in the Local Environment: A Retrospective Cross-Sectional Study. *American Journal of Health Promotion* **28**, e137-e145, doi:10.4278/ajhp.130214-QUAN-70 (2013).
- 378 Poti, J. M., Duffey, K. J. & Popkin, B. M. The association of fast food consumption with poor dietary outcomes and obesity among children: is it the fast food or the remainder of the diet? *Am J Clin Nutr* **99**, 162-171, doi:10.3945/ajcn.113.071928 (2014).
- 379 Biehl, A. *et al.* Parental marital status and childhood overweight and obesity in Norway: a nationally representative cross-sectional study. *BMJ open* **4**, e004502, doi:10.1136/bmjopen-2013-004502 (2014).
- 380 Gallos, L. K., Barttfeld, P., Havlin, S., Sigman, M. & Makse, H. A. Collective behavior in the spatial spreading of obesity. *Scientific reports* **2**, 454, doi:10.1038/srep00454 (2012).
- 381 Popkin, B. M. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. *Am J Clin Nutr* **84**, 289-298 (2006).
- 382 Portion distortion How much are we really eating? , (British Heart Foundation, 2013).
- 383 Wansink, B. & Chandon, P. Meal size, not body size, explains errors in estimating the calorie content of meals. *Annals of internal medicine* **145**, 326-332 (2006).
- 384 Rolls, B. J. The Supersizing of America: Portion Size and the Obesity Epidemic. *Nutrition today* **38**, 42-53 (2003).
- 385 Rolls, B. J., Engell, D. & Birch, L. L. Serving portion size influences 5-year-old but not 3-year-old children's food intakes. *Journal of the American Dietetic Association* **100**, 232-234, doi:10.1016/S0002-8223(00)00070-5 (2000).



- 386 Kemps, E., Tiggemann, M. & Hollitt, S. Exposure to television food advertising primes food-related cognitions and triggers motivation to eat. *Psychology & health* **29**, 1192-1205, doi:10.1080/08870446.2014.918267 (2014).
- 387 Ahima, R. S. The end of overeating: Taking control of the insatiable American appetite. *J. Clin. Invest.* **119**, 2867-2867, doi:10.1172/JCI40983 (2009).
- 388 Organization, W. H. Obesity: Preventing and Managing the Global Epidemic. (Geneva: World Health Organization, 2000).
- 389 Prentice, A. M. & Jebb, S. A. Obesity in Britain: gluttony or sloth? *BMJ* **311**, 437-439 (1995).
- 390 Schiffman, S. S. Appetite and Body Weight Regulation. Sugar, Fat, and Macronutrient Substitutes. *Am. J. Clin. Nutr.* **60**, 644 (1994).
- 391 Vartanian, L. R., Schwartz, M. B. & Brownell, K. D. Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis. *American journal of public health* **97**, 667-675, doi:10.2105/AJPH.2005.083782 (2007).
- 392 Zhen, C., Finkelstein, E. A., Nonnemaker, J., Karns, S. & Todd, J. E. Predicting the Effects of Sugar-Sweetened Beverage Taxes on Food and Beverage Demand in a Large Demand System. *Am J Agric Econ* **96**, 1-25, doi:10.1093/ajae/aat049 (2014).
- 393 Briggs, A. D. *et al.* Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in UK: econometric and comparative risk assessment modelling study. *BMJ* **347**, f6189, doi:10.1136/bmj.f6189 (2013).
- 394 Thulier, D. Breastfeeding in America: a history of influencing factors. *Journal of human lactation : official journal of International Lactation Consultant Association* **25**, 85-94, doi:10.1177/0890334408324452 (2009).
- 395 Butte, N. F. & King, J. C. Energy requirements during pregnancy and lactation. *Public health nutrition* **8**, 1010-1027 (2005).
- 396 Hilson, J. A., Rasmussen, K. M. & Kjolhede, C. L. High prepregnant body mass index is associated with poor lactation outcomes among white, rural women independent of psychosocial and demographic correlates. *Journal of human lactation : official journal of International Lactation Consultant Association* **20**, 18-29 (2004).
- 397 Rasmussen, K. M. & Kjolhede, C. L. Prepregnant overweight and obesity diminish the prolactin response to suckling in the first week postpartum. *Pediatrics* **113**, e465-471 (2004).
- 398 Harder, T., Bergmann, R., Kallischnigg, G. & Plagemann, A. Duration of breastfeeding and risk of overweight: a meta-analysis. *American journal of epidemiology* **162**, 397-403, doi:10.1093/aje/kwi222 (2005).
- 399 Peneau, S., Hercberg, S. & Rolland-Cachera, M. F. Breastfeeding, early nutrition, and adult body fat. *J Pediatr* **164**, 1363-1368, doi:10.1016/j.jpeds.2014.02.020 (2014).
- 400 Jost, T., Lacroix, C., Braegger, C. P., Rochat, F. & Chassard, C. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environmental microbiology*, doi:10.1111/1462-2920.12238 (2013).
- 401 Bergstrom, A. *et al.* Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. *Applied and environmental microbiology* **80**, 2889-2900, doi:10.1128/AEM.00342-14 (2014).
- 402 Wilding, J. Are the causes of obesity primarily environmental? Yes. *BMJ* **345**, e5843, doi:10.1136/bmj.e5843 (2012).
- 403 Frayling, T. M. Are the causes of obesity primarily environmental? No. *BMJ* **345**, e5844, doi:10.1136/bmj.e5844 (2012).
- 404 Turula, M., Kaprio, J., Rissanen, A. & Koskenvuo, M. Body weight in the Finnish Twin Cohort. *Diabetes Res Clin Pract* **10 Suppl 1**, S33-36 (1990).
- 405 Stunkard, A. J., Harris, J. R., Pedersen, N. L. & McClearn, G. E. The body-mass index of twins who have been reared apart. *N Engl J Med* **322**, 1483-1487, doi:10.1056/NEJM199005243222102 (1990).
- 406 Stunkard, A. J., Foch, T. T. & Hrubec, Z. A twin study of human obesity. *JAMA : the journal of the American Medical Association* **256**, 51-54 (1986).

407 Stunkard, A. J. *et al.* An adoption study of human obesity. *N Engl J Med* **314**, 193-198,  
doi:10.1056/NEJM198601233140401 (1986).

408 Rankinen, T. *et al.* The human obesity gene map: the 2005 update. *Obesity* **14**, 529-644,  
doi:10.1038/oby.2006.71 (2006).

409 Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated  
with body mass index. *Nat Genet* **42**, 937-948, doi:10.1038/ng.686 (2010).

410 Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and  
reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949-960,  
doi:10.1038/ng.685 (2010).

411 Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000  
individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* **9**,  
e1003500, doi:10.1371/journal.pgen.1003500 (2013).

412 Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric  
traits and provides insights into genetic architecture. *Nat Genet* **45**, 501-512,  
doi:10.1038/ng.2606 (2013).

413 Bradfield, J. P. *et al.* A genome-wide association meta-analysis identifies new childhood  
obesity loci. *Nat Genet* **44**, 526-531, doi:10.1038/ng.2247 (2012).

414 Wen, W. *et al.* Meta-analysis identifies common variants associated with body mass index in  
east Asians. *Nat Genet* **44**, 307-311, doi:10.1038/ng.1087 (2012).

415 Okada, Y. *et al.* Common variants at CDKAL1 and KLF9 are associated with body mass index in  
east Asian populations. *Nat Genet* **44**, 302-306, doi:10.1038/ng.1086 (2012).

416 Wen, W. *et al.* Meta-analysis of genome-wide association studies in East Asian-ancestry  
populations identifies four new loci for body mass index. *Hum Mol Genet*,  
doi:10.1093/hmg/ddu248 (2014).

417 Walters, R. G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome  
16p11.2. *Nature* **463**, 671-675, doi:10.1038/nature08727 (2010).

418 Glessner, J. T. *et al.* A genome-wide study reveals copy number variants exclusive to  
childhood obesity cases. *Am J Hum Genet* **87**, 661-666, doi:10.1016/j.ajhg.2010.09.014  
(2010).

419 Fischer, J. *et al.* Inactivation of the Fto gene protects from obesity. *Nature* **458**, 894-898,  
doi:10.1038/nature07848 (2009).

420 Church, C. *et al.* Overexpression of Fto leads to increased food intake and results in obesity.  
*Nat Genet* **42**, 1086-1092, doi:10.1038/ng.713 (2010).

421 Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional  
connections with IRX3. *Nature* **507**, 371-375, doi:10.1038/nature13138 (2014).

422 Falchi, M. *et al.* Low copy number of the salivary amylase gene predisposes to obesity. *Nat  
Genet* **46**, 492-497, doi:10.1038/ng.2939 (2014).

423 Marques-Bonet, T. *et al.* A burst of segmental duplications in the genome of the African great  
ape ancestor. *Nature* **457**, 877-881, doi:10.1038/nature07744 (2009).

424 Karlsson, C. L. *et al.* The microbiota of the gut in preschool children with normal and  
excessive body weight. *Obesity* **20**, 2257-2261, doi:10.1038/oby.2012.110 (2012).

425 Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to  
high-fat, high-sucrose diet in mice. *Cell Metab* **17**, 141-152, doi:10.1016/j.cmet.2012.12.007  
(2013).

426 Jackson, R. S. *et al.* Obesity and impaired prohormone processing associated with mutations  
in the human prohormone convertase 1 gene. *Nat Genet* **16**, 303-306, doi:10.1038/ng0797-  
303 (1997).

427 Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early-onset  
obesity in humans. *Nature* **387**, 903-908, doi:10.1038/43185 (1997).

428 Farooqi, I. S. & O'Rahilly, S. Monogenic obesity in humans. *Annual review of medicine* **56**,  
443-458, doi:10.1146/annurev.med.56.062904.144924 (2005).

- 429 Farooqi, I. S. & O'Rahilly, S. Mutations in ligands and receptors of the leptin-melanocortin pathway that lead to obesity. *Nature clinical practice. Endocrinology & metabolism* **4**, 569-577, doi:10.1038/ncpendmet0966 (2008).
- 430 O'Rahilly, S. & Farooqi, I. S. Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions. *Diabetes* **57**, 2905-2910, doi:10.2337/db08-0210 (2008).
- 431 Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398-401, doi:10.1038/32911 (1998).
- 432 Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet* **19**, 155-157, doi:10.1038/509 (1998).
- 433 Clement, K. *et al.* Unexpected endocrine features and normal pigmentation in a young adult patient carrying a novel homozygous mutation in the POMC gene. *J Clin Endocrinol Metab* **93**, 4955-4962, doi:10.1210/jc.2008-1164 (2008).
- 434 Jackson, R. S. *et al.* Small-intestinal dysfunction accompanies the complex endocrinopathy of human proprotein convertase 1 deficiency. *J Clin Invest* **112**, 1550-1560, doi:10.1172/JCI18784 (2003).
- 435 Farooqi, I. S. *et al.* Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**, 1085-1095, doi:10.1056/NEJMoa022050 (2003).
- 436 Stutzmann, F. *et al.* Prevalence of melanocortin-4 receptor deficiency in Europeans and their age-dependent penetrance in multigenerational pedigrees. *Diabetes* **57**, 2511-2518, doi:10.2337/db08-0153 (2008).
- 437 Farooqi, I. S. *et al.* Heterozygosity for a POMC-null mutation and increased obesity risk in humans. *Diabetes* **55**, 2549-2553, doi:10.2337/db06-0214 (2006).
- 438 Creemers, J. W. *et al.* Heterozygous mutations causing partial prohormone convertase 1 deficiency contribute to human obesity. *Diabetes* **61**, 383-390, doi:10.2337/db11-0305 (2012).
- 439 Farooqi, I. S. *et al.* Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med* **356**, 237-247, doi:10.1056/NEJMoa063988 (2007).
- 440 Farooqi, I. S. *et al.* Partial leptin deficiency and human adiposity. *Nature* **414**, 34-35, doi:10.1038/35102112 (2001).
- 441 Han, J. C. *et al.* Brain-derived neurotrophic factor and obesity in the WAGR syndrome. *N Engl J Med* **359**, 918-927, doi:10.1056/NEJMoa0801119 (2008).
- 442 Shalata, A. *et al.* Morbid obesity resulting from inactivation of the ciliary protein CEP19 in humans and mice. *Am J Hum Genet* **93**, 1061-1071, doi:10.1016/j.ajhg.2013.10.025 (2013).
- 443 Keramati, A. R. *et al.* A form of the metabolic syndrome associated with mutations in DYRK1B. *N Engl J Med* **370**, 1909-1919, doi:10.1056/NEJMoa1301824 (2014).
- 444 Pearce, L. R. *et al.* KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* **155**, 765-777, doi:10.1016/j.cell.2013.09.058 (2013).
- 445 Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat Genet* **20**, 113-114, doi:10.1038/2407 (1998).
- 446 Yeo, G. S. *et al.* A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat Genet* **20**, 111-112, doi:10.1038/2404 (1998).
- 447 Asai, M. *et al.* Loss of function of the melanocortin 2 receptor accessory protein 2 is associated with mammalian obesity. *Science* **341**, 275-278, doi:10.1126/science.1233000 (2013).
- 448 Yeo, G. S. *et al.* A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. *Nat Neurosci* **7**, 1187-1189, doi:10.1038/nn1336 (2004).
- 449 Doche, M. E. *et al.* Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest* **122**, 4732-4736, doi:10.1172/JCI62696 (2012).
- 450 Bonnefond, A. *et al.* Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features. *J Clin Invest* **123**, 3037-3041, doi:10.1172/JCI68035 (2013).

451 Ramachandrappa, S. *et al.* Rare variants in single-minded 1 (SIM1) are associated with severe  
obesity. *J Clin Invest* **123**, 3042-3050, doi:10.1172/JCI68016 (2013).

452 Holder, J. L., Jr., Butte, N. F. & Zinn, A. R. Profound obesity associated with a balanced  
translocation that disrupts the SIM1 gene. *Hum Mol Genet* **9**, 101-108 (2000).

453 Gray, J. *et al.* Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity  
associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF)  
gene. *Diabetes* **55**, 3366-3371, doi:10.2337/db06-0550 (2006).

454 Bonaglia, M. C. *et al.* Detailed phenotype-genotype study in five patients with chromosome  
6q16 deletion: narrowing the critical region for Prader-Willi-like phenotype. *European journal  
of human genetics : EJHG* **16**, 1443-1449, doi:10.1038/ejhg.2008.119 (2008).

455 Kernie, S. G., Liebl, D. J. & Parada, L. F. BDNF regulates eating behavior and locomotor activity  
in mice. *The EMBO journal* **19**, 1290-1300, doi:10.1093/emboj/19.6.1290 (2000).

456 Michaud, J. L. *et al.* Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the  
paraventricular nucleus of the hypothalamus. *Hum Mol Genet* **10**, 1465-1473 (2001).

457 Xu, B. *et al.* Brain-derived neurotrophic factor regulates energy balance downstream of  
melanocortin-4 receptor. *Nat Neurosci* **6**, 736-742, doi:10.1038/nn1073 (2003).

458 Silos-Santiago, I., Fagan, A. M., Garber, M., Fritzsche, B. & Barbacid, M. Severe sensory deficits  
but normal CNS development in newborn mice lacking TrkB and TrkC tyrosine protein kinase  
receptors. *The European journal of neuroscience* **9**, 2045-2056 (1997).

459 Sebag, J. A., Zhang, C., Hinkle, P. M., Bradshaw, A. M. & Cone, R. D. Developmental control of  
the melanocortin-4 receptor by MRAP2 proteins in zebrafish. *Science* **341**, 278-281,  
doi:10.1126/science.1232995 (2013).

460 Hearn, T. *et al.* Mutation of ALMS1, a large gene with a tandem repeat encoding 47 amino  
acids, causes Alstrom syndrome. *Nat Genet* **31**, 79-83, doi:10.1038/ng874 (2002).

461 Collin, G. B. *et al.* Mutations in ALMS1 cause obesity, type 2 diabetes and neurosensory  
degeneration in Alstrom syndrome. *Nat Genet* **31**, 74-78, doi:10.1038/ng867 (2002).

462 Myktytn, K. *et al.* Identification of the gene (BBS1) most commonly involved in Bardet-Biedl  
syndrome, a complex human obesity syndrome. *Nat Genet* **31**, 435-438, doi:10.1038/ng935  
(2002).

463 Patten, J. L. *et al.* Mutation in the gene encoding the stimulatory G protein of adenylate  
cyclase in Albright's hereditary osteodystrophy. *N Engl J Med* **322**, 1412-1419,  
doi:10.1056/NEJM199005173222002 (1990).

464 Borman, A. D. *et al.* A homozygous mutation in the TUB gene associated with retinal  
dystrophy and obesity. *Hum Mutat* **35**, 289-293, doi:10.1002/humu.22482 (2014).

465 Kolehmainen, J. *et al.* Cohen syndrome is caused by mutations in a novel gene, COH1,  
encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and  
intracellular protein transport. *Am J Hum Genet* **72**, 1359-1369 (2003).

466 Butler, M. G., Meaney, F. J. & Palmer, C. G. Clinical and cytogenetic survey of 39 individuals  
with Prader-Labhart-Willi syndrome. *American journal of medical genetics* **23**, 793-809,  
doi:10.1002/ajmg.1320230307 (1986).

467 Schaaf, C. P. *et al.* Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and  
autism. *Nat Genet* **45**, 1405-1408, doi:10.1038/ng.2776 (2013).

468 Stelzer, Y., Sagi, I., Yanuka, O., Eiges, R. & Benvenisty, N. The noncoding RNA IPW regulates  
the imprinted DLK1-DIO3 locus in an induced pluripotent stem cell model of Prader-Willi  
syndrome. *Nat Genet* **46**, 551-557, doi:10.1038/ng.2968 (2014).

469 Ansley, S. J. *et al.* Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl  
syndrome. *Nature* **425**, 628-633, doi:10.1038/nature02030 (2003).

470 Seo, S. *et al.* Requirement of Bardet-Biedl syndrome proteins for leptin receptor signaling.  
*Hum Mol Genet* **18**, 1323-1331, doi:10.1093/hmg/ddp031 (2009).

471 Chen, M. *et al.* Central nervous system imprinting of the G protein G(s)alpha and its role in  
metabolic regulation. *Cell Metab* **9**, 548-555, doi:10.1016/j.cmet.2009.05.004 (2009).



472 Chandler, K. E. *et al.* Diagnostic criteria, clinical characteristics, and natural history of Cohen  
syndrome. *Journal of medical genetics* **40**, 233-241 (2003).

473 Prada, P. O. *et al.* Tub has a key role in insulin and leptin signaling and action in vivo in  
hypothalamic nuclei. *Diabetes* **62**, 137-148, doi:10.2337/db11-1388 (2013).

474 Dawn Teare, M. & Barrett, J. H. Genetic linkage studies. *The Lancet* **366**, 1036-1044 (2005).

475 Li, M., Boehnke, M. & Abecasis, G. R. Efficient study designs for test of genetic association  
using sibship data and unrelated cases and controls. *Am J Hum Genet* **78**, 778-792,  
doi:10.1086/503711 (2006).

476 Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association  
studies. *Nature reviews. Genetics* **12**, 465-474, doi:10.1038/nrg2989 (2011).

477 Morton, N. E. Sequential tests for the detection of linkage. *Am J Hum Genet* **7**, 277-318  
(1955).

478 Fiskerstrand, T. *et al.* Familial diarrhea syndrome caused by an activating GUCY2C mutation.  
*N Engl J Med* **366**, 1586-1595, doi:10.1056/NEJMoa1110132 (2012).

479 Hu, H. *et al.* A unified test of linkage analysis and rare-variant association for analysis of  
pedigree sequence data. *Nature biotechnology* **32**, 663-669, doi:10.1038/nbt.2895 (2014).

480 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors.  
*Proceedings of the National Academy of Sciences* **74**, 5463-5467 (1977).

481 Sanger, F. *et al.* Nucleotide sequence of bacteriophage [phi]X174 DNA. *Nature* **265**, 687-695  
(1977).

482 Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74**,  
560-564 (1977).

483 Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed  
synthesis with DNA polymerase. *Journal of molecular biology* **94**, 441-448 (1975).

484 Singleton, A. B. Exome sequencing: a transformative technology. *Lancet neurology* **10**, 942-  
946, doi:10.1016/S1474-4422(11)70196-X (2011).

485 Brouzes, E. *et al.* Droplet microfluidic technology for single-cell high-throughput screening.  
*Proceedings of the National Academy of Sciences* **106**, 14195-14200,  
doi:10.1073/pnas.0903542106 (2009).

486 Tewhey, R. *et al.* Microdroplet-based PCR enrichment for large-scale targeted sequencing.  
*Nature biotechnology* **27**, 1025-1031, doi:10.1038/nbt.1583 (2009).

487 Illumina, I. *Illumina website*, <<http://www.illumina.com/>> (2014).

488 Dai, M. *et al.* NGSQC: cross-platform quality analysis pipeline for deep sequencing data. *BMC  
genomics* **11 Suppl 4**, S7, doi:10.1186/1471-2164-11-S4-S7 (2010).

489 Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment of Illumina  
second-generation sequencing data. *BMC Bioinformatics* **11**, 485, doi:10.1186/1471-2105-11-  
485 (2010).

490 Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read  
data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**, e105,  
doi:10.1093/nar/gkn425 (2008).

491 Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-  
generation sequencing data. *Nature reviews. Genetics* **12**, 443-451, doi:10.1038/nrg2986  
(2011).

492 Raney, B. J. *et al.* ENCODE whole-genome data in the UCSC genome browser (2011 update).  
*Nucleic Acids Res* **39**, D871-875, doi:10.1093/nar/gkq1017 (2011).

493 The Genome Reference Consortium,  
<<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/>> (

494 Pabinger, S. *et al.* A survey of tools for variant analysis of next-generation genome  
sequencing data. *Briefings in bioinformatics* **15**, 256-278, doi:10.1093/bib/bbs086 (2014).

495 Neuman, J. A., Isakov, O. & Shomron, N. Analysis of insertion-deletion from deep-sequencing  
data: software evaluation for optimal detection. *Briefings in bioinformatics* **14**, 46-55,  
doi:10.1093/bib/bbs013 (2013).

- 496 Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a  
wealth of genomic data. *Nature reviews. Genetics* **12**, 628-640, doi:10.1038/nrg3046 (2011).
- 497 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification  
strategies for exome sequencing. *European journal of human genetics : EJHG* **20**, 490-497,  
doi:10.1038/ejhg.2011.258 (2012).
- 498 Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery.  
*Nature reviews. Genetics* **12**, 745-755, doi:10.1038/nrg3031 (2011).
- 499 Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad  
Sci U S A* **107**, 961-968, doi:10.1073/pnas.0912629107 (2010).
- 500 Mathe, E. *et al.* Computational approaches for predicting the biological effect of p53  
missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids  
Res* **34**, 1317-1325, doi:10.1093/nar/gkj518 (2006).
- 501 Wei, Q., Wang, L., Wang, Q., Kruger, W. D. & Dunbrack, R. L., Jr. Testing computational  
prediction of missense mutation phenotypes: functional characterization of 204 mutations of  
human cystathionine beta synthase. *Proteins* **78**, 2058-2074, doi:10.1002/prot.22722 (2010).
- 502 Lindblom, A. & Robinson, P. N. Bioinformatics for human genetics: promises and challenges.  
*Hum Mutat* **32**, 495-500, doi:10.1002/humu.21468 (2011).
- 503 Hodges, E. *et al.* Genome-wide in situ exon capture for selective resequencing. *Nat Genet* **39**,  
1522-1527, doi:10.1038/ng.2007.42 (2007).
- 504 Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat  
Methods* **4**, 903-905, doi:10.1038/nmeth1111 (2007).
- 505 Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA  
sequencing. *Proc Natl Acad Sci U S A* **106**, 19096-19101, doi:10.1073/pnas.0910672106  
(2009).
- 506 Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes.  
*Nature* **461**, 272-276, doi:10.1038/nature08250 (2009).
- 507 Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**,  
30-35, doi:10.1038/ng.499 (2010).
- 508 Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.  
*Nat Genet* **42**, 790-793, doi:10.1038/ng.646 (2010).
- 509 Girard, C. A. *et al.* Functional analysis of six Kir6.2 (KCNJ11) mutations causing neonatal  
diabetes. *Pflugers Archiv : European journal of physiology* **453**, 323-332, doi:10.1007/s00424-  
006-0112-3 (2006).
- 510 Pearson, E. R. *et al.* Switching from insulin to oral sulfonylureas in patients with diabetes due  
to Kir6.2 mutations. *N Engl J Med* **355**, 467-477, doi:10.1056/NEJMoa061759 (2006).
- 511 Aquilante, C. L. Sulfonylurea pharmacogenomics in Type 2 diabetes: the influence of drug  
target and diabetes risk polymorphisms. *Expert review of cardiovascular therapy* **8**, 359-372,  
doi:10.1586/erc.09.154 (2010).
- 512 Pearson, E. R. *et al.* Genetic cause of hyperglycaemia and response to treatment in diabetes.  
*Lancet* **362**, 1275-1281, doi:10.1016/S0140-6736(03)14571-0 (2003).
- 513 Tarasov, A. I. *et al.* A rare mutation in ABCC8/SUR1 leading to altered ATP-sensitive K<sup>+</sup>  
channel activity and beta-cell glucose sensing is associated with type 2 diabetes in adults.  
*Diabetes* **57**, 1595-1604, doi:10.2337/db07-1547 (2008).
- 514 Bowman, P. *et al.* Heterozygous ABCC8 mutations are a cause of MODY. *Diabetologia* **55**,  
123-127, doi:10.1007/s00125-011-2319-x (2012).
- 515 Riveline, J. P. *et al.* Clinical and metabolic features of adult-onset diabetes caused by ABCC8  
mutations. *Diabetes Care* **35**, 248-251, doi:10.2337/dc11-1469 (2012).
- 516 Gloyn, A. L. *et al.* Activating mutations in the gene encoding the ATP-sensitive potassium-  
channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* **350**, 1838-1849,  
doi:10.1056/NEJMoa032922 (2004).
- 517 Hennekam, R. C. & Biesecker, L. G. Next-generation sequencing demands next-generation  
phenotyping. *Hum Mutat* **33**, 884-886, doi:10.1002/humu.22048 (2012).

518 Del Fabbro, C., Scalabrin, S., Morgante, M. & Giorgi, F. M. An extensive evaluation of read  
trimming effects on Illumina NGS data analysis. *PLoS One* **8**, e85024,  
doi:10.1371/journal.pone.0085024 (2013).

519 O'Rawe, J. *et al.* Low concordance of multiple variant-calling pipelines: practical implications  
for exome and genome sequencing. *Genome Med* **5**, 28, doi:10.1186/gm432 (2013).

520 McCarthy, D. *et al.* Choice of transcripts and software has a large effect on variant  
annotation. *Genome Medicine* **6**, 26 (2014).

521 Shigemizu, D. *et al.* A practical method to detect SNVs and indels from whole genome and  
exome sequencing data. *Scientific reports* **3**, 2161, doi:10.1038/srep02161 (2013).

522 Greeley, S. A. *et al.* The cost-effectiveness of personalized genetic medicine: the case of  
genetic testing in neonatal diabetes. *Diabetes Care* **34**, 622-627, doi:10.2337/dc10-1616  
(2011).

523 Bonnefond, A. *et al.* Molecular diagnosis of neonatal diabetes mellitus using next-generation  
sequencing of the whole exome. *PLoS One* **5**, e13630, doi:10.1371/journal.pone.0013630  
(2010).

524 Green, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical  
exome and genome sequencing. *Genetics in medicine : official journal of the American  
College of Medical Genetics* **15**, 565-574, doi:10.1038/gim.2013.73 (2013).

525 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform.  
*Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

526 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and  
human phenotype. *Nucleic Acids Res* **42**, D980-985, doi:10.1093/nar/gkt1113 (2014).

527 Stenson, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive  
mutation repository for clinical and molecular genetics, diagnostic testing and personalized  
genomic medicine. *Human genetics* **133**, 1-9, doi:10.1007/s00439-013-1358-4 (2014).

528 Shields, B. M. *et al.* Maturity-onset diabetes of the young (MODY): how many cases are we  
missing? *Diabetologia* **53**, 2504-2508, doi:10.1007/s00125-010-1799-4 (2010).

529 Naylor, R. N. *et al.* Cost-effectiveness of MODY genetic testing: translating genomic advances  
into practical health applications. *Diabetes Care* **37**, 202-209, doi:10.2337/dc13-0410 (2014).

530 Creemers, J. W. M. *et al.* Heterozygous mutations causing partial prohormone convertase 1  
deficiency contribute to human obesity. *Diabetes* **61**, 383-390, doi:10.2337/db11-0305  
(2012).

531 Dubern, B. *et al.* Mutational analysis of the pro-opiomelanocortin gene in French obese  
children led to the identification of a novel deleterious heterozygous mutation located in the  
alpha-melanocyte stimulating hormone domain. *Pediatric research* **63**, 211-216,  
doi:10.1203/PDR.0b013e31815ed62b (2008).

532 Farooqi, I. S. *et al.* Hyperphagia and early-onset obesity due to a novel homozygous missense  
mutation in prohormone convertase 1/3. *J Clin Endocrinol Metab* **92**, 3369-3373,  
doi:10.1210/jc.2007-0687 (2007).

533 Frank, G. R. *et al.* Severe obesity and diabetes insipidus in a patient with PCSK1 deficiency.  
*Mol Genet Metab* **110**, 191-194, doi:10.1016/j.ymgme.2013.04.005 (2013).

534 Martin, M. G. *et al.* Congenital proprotein convertase 1/3 deficiency causes malabsorptive  
diarrhea and other endocrinopathies in a pediatric cohort. *Gastroenterology* **145**, 138-148,  
doi:10.1053/j.gastro.2013.03.048 (2013).

535 Goodman, L. J. & Gorman, C. M. Autoproteolytic activation of the mouse prohormone  
convertase mPC1. *Biochem Biophys Res Commun* **201**, 795-804, doi:10.1006/bbrc.1994.1771  
(1994).

536 Rabah, N., Gauthier, D., Wilkes, B. C., Gauthier, D. J. & Lazure, C. Single amino acid  
substitution in the PC1/3 propeptide can induce significant modifications of its inhibitory  
profile toward its cognate enzyme. *J Biol Chem* **281**, 7556-7567,  
doi:10.1074/jbc.M510607200 (2006).



- 537 Boudreault, A., Gauthier, D. & Lazure, C. Proprotein convertase PC1/3-related peptides are  
potent slow tight-binding inhibitors of murine PC1/3 and Hfurin. *J Biol Chem* **273**, 31574-  
31580 (1998).
- 538 Seidah, N. G. & Prat, A. The proprotein convertases are potential targets in the treatment of  
dyslipidemia. *Journal of molecular medicine* **85**, 685-696, doi:10.1007/s00109-007-0172-7  
(2007).
- 539 Benzinou, M. *et al.* Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nat*  
*Genet* **40**, 943-945, doi:10.1038/ng.177 (2008).
- 540 Strawbridge, R. J. *et al.* Genome-wide association identifies nine common variants associated  
with fasting proinsulin levels and provides new insights into the pathophysiology of type 2  
diabetes. *Diabetes* **60**, 2624-2634, doi:10.2337/db11-0415 (2011).
- 541 Zimmer, C. in *The New York Times* (2014).
- 542 Wilson, M. R. *et al.* Actionable diagnosis of neuroleptospirosis by next-generation  
sequencing. *N Engl J Med* **370**, 2408-2417, doi:10.1056/NEJMoa1401268 (2014).
- 543 Need, A. C. *et al.* Clinical application of exome sequencing in undiagnosed genetic conditions.  
*Journal of medical genetics* **49**, 353-361, doi:10.1136/jmedgenet-2012-100819 (2012).
- 544 Enns, G. M. *et al.* Mutations in NGLY1 cause an inherited disorder of the endoplasmic  
reticulum-associated degradation pathway. *Genetics in medicine : official journal of the*  
*American College of Medical Genetics*, doi:10.1038/gim.2014.22 (2014).
- 545 *Hunting down my son's killer* <<http://matt.might.net/articles/my-sons-killer/>> (2012).
- 546 Atwal, P. S. *et al.* Clinical whole-exome sequencing: are we there yet? *Genetics in medicine :*  
*official journal of the American College of Medical Genetics*, doi:10.1038/gim.2014.10 (2014).
- 547 Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes.  
*Cell* **148**, 1293-1307, doi:10.1016/j.cell.2012.02.009 (2012).
- 548 Wu, Y. *et al.* Multilayered genetic and omics dissection of mitochondrial activity in a mouse  
reference population. *Cell* **158**, 1415-1430, doi:10.1016/j.cell.2014.07.039 (2014).
- 549 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**,  
75-82, doi:10.1038/nature11232 (2012).
- 550 Consortium, E. P. *et al.* An integrated encyclopedia of DNA elements in the human genome.  
*Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
- 551 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101-108,  
doi:10.1038/nature11233 (2012).
- 552 Gaur, D. *et al.* On the immortality of television sets: "function" in the human genome  
according to the evolution-free gospel of ENCODE. *Genome biology and evolution* **5**, 578-590,  
doi:10.1093/gbe/evt028 (2013).
- 553 Rands, C. M., Meader, S., Ponting, C. P. & Lunter, G. 8.2% of the Human Genome Is  
Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human  
Lineage. *PLoS Genet* **10**, e1004525, doi:10.1371/journal.pgen.1004525 (2014).
- 554 Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575-581,  
doi:10.1038/nature13302 (2014).
- 555 Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-  
587, doi:10.1038/nature13319 (2014).
- 556 Ezkurdia, I. *et al.* Multiple evidence strands suggest that there may be as few as 19 000  
human protein-coding genes. *Hum Mol Genet*, doi:10.1093/hmg/ddu309 (2014).
- 557 Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**,  
543-550, doi:10.1038/ng.2982 (2014).
- 558 Chauveau, C., Rowell, J. & Ferreira, A. A Rising Titan: TTN Review and Mutation Update. *Hum*  
*Mutat*, n/a-n/a, doi:10.1002/humu.22611 (2014).
- 559 Hu, Y. *et al.* Genotyping performance between saliva and blood-derived genomic DNAs on  
the DMET array: a comparison. *PLoS One* **7**, e33968, doi:10.1371/journal.pone.0033968  
(2012).

- 560 Saunders, C. J. *et al.* Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med* **4**, 154ra135, doi:10.1126/scitranslmed.3004041 (2012).
- 561 Lawrence, L. *et al.* The implications of familial incidental findings from exome sequencing: the NIH Undiagnosed Diseases Program experience. *Genetics in medicine : official journal of the American College of Medical Genetics*, doi:10.1038/gim.2014.29 (2014).
- 562 Yu, J. H., Harrell, T. M., Jamal, S. M., Tabor, H. K. & Bamshad, M. J. Attitudes of genetics professionals toward the return of incidental results from exome and whole-genome sequencing. *Am J Hum Genet* **95**, 77-84, doi:10.1016/j.ajhg.2014.06.004 (2014).
- 563 Schrauwen, I. *et al.* A sensitive and specific diagnostic test for hearing loss using a microdroplet PCR-based approach and next generation sequencing. *American journal of medical genetics. Part A* **161A**, 145-152, doi:10.1002/ajmg.a.35737 (2013).
- 564 Gui, H. *et al.* Targeted Next-Generation Sequencing on Hirschsprung Disease: A Pilot Study Exploits DNA Pooling. *Annals of human genetics*, doi:10.1111/ahg.12076 (2014).
- 565 Harismendy, O. *et al.* Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome biology* **12**, R124, doi:10.1186/gb-2011-12-12-r124 (2011).
- 566 Dames, S. *et al.* The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders. *The Journal of molecular diagnostics : JMD* **15**, 526-534, doi:10.1016/j.jmoldx.2013.03.005 (2013).
- 567 Valencia, C. A. *et al.* Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS One* **8**, e53083, doi:10.1371/journal.pone.0053083 (2013).
- 568 Saeed, S. *et al.* Novel LEPR mutations in obese Pakistani children identified by PCR-based enrichment and next generation sequencing. *Obesity* **22**, 1112-1117, doi:10.1002/oby.20667 (2014).
- 569 Antonietta Robino, N. P., C. Mansfield, D. Hwang, D.R. Reed, P. Gasparini. in *ESHG*.
- 570 Mejean, C. *et al.* Association between intake of nutrients and food groups and liking for fat (The Nutrinet-Sante Study). *Appetite* **78**, 147-155, doi:10.1016/j.appet.2014.03.017 (2014).
- 571 Boesveldt, S. & Lundstrom, J. N. Detecting fat content of food from a distance: olfactory-based fat discrimination in humans. *PLoS One* **9**, e85977, doi:10.1371/journal.pone.0085977 (2014).
- 572 Thiebaud, N. *et al.* Hyperlipidemic diet causes loss of olfactory sensory neurons, reduces olfactory discrimination, and disrupts odor-reversal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 6970-6984, doi:10.1523/JNEUROSCI.3366-13.2014 (2014).
- 573 Gao, H. *et al.* Early B cell factor 1 regulates adipocyte morphology and lipolysis in white adipose tissue. *Cell Metab* **19**, 981-992, doi:10.1016/j.cmet.2014.03.032 (2014).
- 574 Lidell, M. E. *et al.* Evidence for two types of brown adipose tissue in humans. *Nat Med* **19**, 631-634, doi:10.1038/nm.3017 (2013).
- 575 Bostrom, P. *et al.* A PGC1-alpha-dependent myokine that drives brown-fat-like development of white fat and thermogenesis. *Nature* **481**, 463-468, doi:10.1038/nature10777 (2012).
- 576 Rao, R. R. *et al.* Meteorin-like Is a Hormone that Regulates Immune-Adipose Interactions to Increase Beige Fat Thermogenesis. *Cell* **157**, 1279-1291, doi:10.1016/j.cell.2014.03.065 (2014).
- 577 Qiu, Y. *et al.* Eosinophils and type 2 cytokine signaling in macrophages orchestrate development of functional beige fat. *Cell* **157**, 1292-1308, doi:10.1016/j.cell.2014.03.066 (2014).
- 578 Chondronikola, M. *et al.* Brown Adipose Tissue Improves Whole Body Glucose Homeostasis and Insulin Sensitivity in Humans. *Diabetes*, doi:10.2337/db14-0746 (2014).

- 579 Soccio, R. E., Chen, E. R. & Lazar, M. A. Thiazolidinediones and the Promise of Insulin  
Sensitization in Type 2 Diabetes. *Cell Metab* **20**, 573-591, doi:10.1016/j.cmet.2014.08.005  
(2014).
- 580 Cardamone, M. D. *et al.* GPS2/KDM4A Pioneering Activity Regulates Promoter-Specific  
Recruitment of PPARgamma. *Cell reports* **8**, 163-176, doi:10.1016/j.celrep.2014.05.041  
(2014).
- 581 Finan, B. *et al.* Unimolecular dual incretins maximize metabolic benefits in rodents, monkeys,  
and humans. *Sci Transl Med* **5**, 209ra151, doi:10.1126/scitranslmed.3007218 (2013).
- 582 Lam, Y. Y., Peterson, C. M. & Ravussin, E. Resveratrol vs. calorie restriction: data from rodents  
to humans. *Experimental gerontology* **48**, 1018-1024, doi:10.1016/j.exger.2013.04.005  
(2013).
- 583 Poulsen, M. M. *et al.* High-dose resveratrol supplementation in obese men: an investigator-  
initiated, randomized, placebo-controlled clinical trial of substrate metabolism, insulin  
sensitivity, and body composition. *Diabetes* **62**, 1186-1195, doi:10.2337/db12-0975 (2013).
- 584 Roberts, V. H. *et al.* Beneficial and cautionary outcomes of resveratrol supplementation in  
pregnant nonhuman primates. *FASEB journal : official publication of the Federation of  
American Societies for Experimental Biology* **28**, 2466-2477, doi:10.1096/fj.13-245472 (2014).
- 585 Suh, J. M. *et al.* Endocrinization of FGF1 produces a neomorphic and potent insulin sensitizer.  
*Nature advance online publication*, doi:10.1038/nature13540 (2014).
- 586 Yore, M. M. *et al.* Discovery of a Class of Endogenous Mammalian Lipids with Anti-Diabetic  
and Anti-inflammatory Effects. *Cell* **159**, 318-332, doi:10.1016/j.cell.2014.09.035 (2014).
- 587 Rivera, J. F., Costes, S., Gurlo, T., Glabe, C. G. & Butler, P. C. Autophagy defends pancreatic  
beta cells from human islet amyloid polypeptide-induced toxicity. *J Clin Invest*,  
doi:10.1172/JCI71981 (2014).
- 588 Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers.  
*Nature* **500**, 541-546, doi:10.1038/nature12506 (2013).
- 589 Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* **500**,  
585-588, doi:10.1038/nature12480 (2013).
- 590 Woting, A., Pfeiffer, N., Loh, G., Klaus, S. & Blaut, M. Clostridium ramosum Promotes High-Fat  
Diet-Induced Obesity in Gnotobiotic Mouse Models. *mBio* **5**, doi:10.1128/mBio.01530-14  
(2014).
- 591 Joyce, S. A. *et al.* Regulation of host weight gain and lipid metabolism by bacterial bile acid  
modification in the gut. *Proc Natl Acad Sci U S A* **111**, 7421-7426,  
doi:10.1073/pnas.1323599111 (2014).
- 592 Chen, Z. *et al.* Incorporation of therapeutically modified bacteria into gut microbiota inhibits  
obesity. *J Clin Invest*, doi:10.1172/JCI72517 (2014).
- 593 Pagliuca, F. W. *et al.* Generation of Functional Human Pancreatic beta Cells In Vitro. *Cell* **159**,  
428-439, doi:10.1016/j.cell.2014.09.040 (2014).
- 594 Loomis, E. W. *et al.* Sequencing the unsequenceable: expanded CGG-repeat alleles of the  
fragile X gene. *Genome Res* **23**, 121-128, doi:10.1101/gr.141705.112 (2013).
- 595 Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the  
human transcriptome. *Nature biotechnology* **31**, 1009-1014, doi:10.1038/nbt.2705 (2013).
- 596 Au, K. F. *et al.* Characterization of the human ESC transcriptome by hybrid sequencing. *Proc  
Natl Acad Sci U S A* **110**, E4821-4830, doi:10.1073/pnas.1320101110 (2013).
- 597 Pibiri, E., Holzmeister, P., Lalkens, B., Acuna, G. P. & Tinnefeld, P. Single-molecule positioning  
in zeromode waveguides by DNA origami nanoanchors. *Nano letters* **14**, 3499-3503,  
doi:10.1021/nl501064b (2014).
- 598 Larkin, J., Foquet, M., Turner, S. W., Korch, J. & Wanunu, M. Reversible Positioning of Single  
Molecules inside Zero-Mode Waveguides. *Nano letters*, doi:10.1021/nl503134x (2014).
- 599 Shankla, M. & Aksimentiev, A. Conformational transitions and stop-and-go nanopore  
transport of single-stranded DNA on charged graphene. *Nature communications* **5**, 5171,  
doi:10.1038/ncomms6171 (2014).

- 600 Mikheyev, A. S. & Tin, M. M. A first look at the Oxford Nanopore MinION sequencer. *Molecular ecology resources*, doi:10.1111/1755-0998.12324 (2014).
- 601 Laszlo, A. H. *et al.* Decoding long nanopore sequencing reads of natural DNA. *Nature biotechnology*, doi:10.1038/nbt.2950 (2014).
- 602 Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology* **4**, 265-270, doi:10.1038/nnano.2009.12 (2009).
- 603 Kumar, S. *et al.* PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Scientific reports* **2**, 684, doi:10.1038/srep00684 (2012).
- 604 Phillips, L. S., Ratner, R. E., Buse, J. B. & Kahn, S. E. We can change the natural history of type 2 diabetes. *Diabetes Care* **37**, 2668-2676, doi:10.2337/dc14-0817 (2014).
- 605 Shields, B. M. *et al.* The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes. *Diabetologia* **55**, 1265-1272, doi:10.1007/s00125-011-2418-8 (2012).
- 606 Morandi, A. *et al.* Estimation of newborn risk for child or adolescent obesity: lessons from longitudinal birth cohorts. *PLoS One* **7**, e49919, doi:10.1371/journal.pone.0049919 (2012).
- 607 Wang, T. J. *et al.* 2-Aminoadipic acid is a biomarker for diabetes risk. *J Clin Invest* **123**, 4309-4317, doi:10.1172/JCI64801 (2013).
- 608 Guay, C. & Regazzi, R. Circulating microRNAs as novel biomarkers for diabetes mellitus. *Nat Rev Endocrinol* **9**, 513-521, doi:10.1038/nrendo.2013.86 (2013).
- 609 Jais, A. *et al.* Heme oxygenase-1 drives metaflammation and insulin resistance in mouse and man. *Cell* **158**, 25-40, doi:10.1016/j.cell.2014.04.043 (2014).
- 610 Stallbaumer, T. J. *From Jeopardy to Cancer Prediction- IBM's Watson Makes Strides*, <[http://www.uatrav.com/opinion/article\\_63dfb8fe-402b-11e3-a01a-0019bb30f31a.html](http://www.uatrav.com/opinion/article_63dfb8fe-402b-11e3-a01a-0019bb30f31a.html)> (2013).
- 611 Ng, M. C. & Bowden, D. W. Is genetic testing of value in predicting and treating obesity? *North Carolina medical journal* **74**, 530-533 (2013).
- 612 Vaxillaire, M. *et al.* Type 2 diabetes-related genetic risk scores associated with variations in fasting plasma glucose and development of impaired glucose homeostasis in the prospective DESIR study. *Diabetologia* **57**, 1601-1610, doi:10.1007/s00125-014-3277-x (2014).
- 613 Witte, J. S., Visscher, P. M. & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. *Nature reviews. Genetics* **advance online publication**, doi:10.1038/nrg3786
- <http://www.nature.com/nrg/journal/vaop/ncurrent/abs/nrg3786.html#supplementary-information> (2014).
- 614 Consortium, S. T. D. *et al.* Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA : the journal of the American Medical Association* **311**, 2305-2314, doi:10.1001/jama.2014.6511 (2014).
- 615 Press, M. O., Carlson, K. D. & Queitsch, C. The overdue promise of short tandem repeat variation for heritability. *Trends in genetics : TIG*, doi:10.1016/j.tig.2014.07.008 (2014).
- 616 Rios, J. J. & Delgado, M. R. Using whole-exome sequencing to identify variants inherited from mosaic parents. *European journal of human genetics : EJHG*, doi:10.1038/ejhg.2014.125 (2014).
- 617 Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **509**, 363-369, doi:10.1038/nature13437 (2014).
- 618 Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).
- 619 Jacobs, K. *et al.* Low-grade chromosomal mosaicism in human somatic and embryonic stem cell populations. *Nature communications* **5**, 4227, doi:10.1038/ncomms5227 (2014).
- 620 Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, doi:10.1093/bioinformatics/btu356 (2014).

- 621 Meltz Steinberg, K. S. V. K. G.-L. T. A. F. R. S. A. R. H. J. S. S. A. M. A. S. U. W. W. C. C. D. Single  
haplotype assembly of the human genome from a hydatidiform mole. *bioRxiv* (2014).
- 622 Guo, X. *et al.* Genome reference and sequence variation in the large repetitive central exon  
of human MUC5AC. *American journal of respiratory cell and molecular biology* **50**, 223-232,  
doi:10.1165/rcmb.2013-0235OC (2014).
- 623 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE  
Project. *Genome Res* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).
- 624 Wijaya, E., Frith, M. C., Horton, P. & Asai, K. Finding protein-coding genes through human  
polymorphisms. *PLoS One* **8**, e54210, doi:10.1371/journal.pone.0054210 (2013).
- 625 Reichardt, J. K. Quo vadis, genoma? A call to pipettes for biochemists. *Trends in biochemical  
sciences* **32**, 529-530, doi:10.1016/j.tibs.2007.10.001 (2007).
- 626 Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome.  
*Proc Natl Acad Sci U S A* **104**, 19428-19433, doi:10.1073/pnas.0709013104 (2007).
- 627 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial  
immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
- 628 Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of  
genomic regions by multiplex homology-directed repair. *Nature*, doi:10.1038/nature13695  
(2014).
- 629 Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the  
CRISPR-Cas9 system. *Science* **343**, 80-84, doi:10.1126/science.1246981 (2014).
- 630 Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in  
human cells. *Nature* **509**, 487-491, doi:10.1038/nature13166 (2014).
- 631 Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**,  
84-87, doi:10.1126/science.1247005 (2014).
- 632 MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-  
coding genes. *Science* **335**, 823-828, doi:10.1126/science.1215040 (2012).
- 633 Flannick, J. *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat  
Genet* **46**, 357-363, doi:10.1038/ng.2915 (2014).
- 634 Jacquemont, S. *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the  
chromosome 16p11.2 locus. *Nature* **478**, 97-102, doi:10.1038/nature10406 (2011).
- 635 Neel, J. V. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J  
Hum Genet* **14**, 353-362 (1962).
- 636 Speakman, J. R. Thrifty genes for obesity, an attractive but flawed idea, and an alternative  
perspective: the 'drifty gene' hypothesis. *Int J Obes (Lond)* **32**, 1611-1617,  
doi:10.1038/ijo.2008.161 (2008).
- 637 Sellayah, D., Cagampang, F. R. & Cox, R. D. On the evolutionary origins of obesity: a new  
hypothesis. *Endocrinology* **155**, 1573-1588, doi:10.1210/en.2013-2103 (2014).
- 638 Schiffels, S. & Durbin, R. Inferring human population size and separation history from  
multiple genome sequences. *Nat Genet*, doi:10.1038/ng.3015 (2014).